Vol. 38, Parts 1 and 2 June 1951

BIOMETRIKA

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR
E. S. PEARSON

ASSOCIATE EDITORS
M. G. KENDALL JOHN WISHART

in consultation with
HARALD CRAMÉR R. C. GEARY
J. B. S. HALDANE

Reprinted by offset-litho, 1960

ISSUED BY
THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON
PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

[*Issued* 12 *June* 1951]

UNIVERSITY OF UTAH LIBRARIES

VOLUME 38, PARTS 1 AND 2 JUNE 1951

MAJOR GREENWOOD

1880–1949

Major Greenwood, Professor Emeritus of Epidemiology and Vital Statistics in the University of London, died suddenly at the age of 69 on 5 October 1949 while attending a scientific meeting. Thus was lost to the medical profession one of its outstanding and most influential specialists and teachers. An only child, he was born in 1880 in the East End of London where his father was in family practice. He was educated on the classical side at Merchant Taylors' School and subsequently, in 1898, with a Buxton scholarship, studied medicine in University College and the London Hospital, gaining the conjoint qualification in 1904. He first acted as assistant to his father, but clearly with little interest in this side of the work for, within a matter of months, he forsook clinical medicine for good and, appointed demonstrator in physiology, he began a happy association with Leonard Hill in the London Hospital Medical College. Here, in addition to physiological research, mainly on the effects of pressure changes in man and animals (on which in 1906–7 he published alone or with Hill no less than eight papers, and devoted his Arris and Gale Lectures (1908) to the same topic), he had already revealed his interest in the special field to which his subsequent professional life was mainly devoted. His first published work, in the year of qualification, was a statistical study of the variability and interrelationships of human visceral weights appearing in this journal in 1904. Others in this early period related to statistical aspects of the opsonic index and the problem of marital infection in pulmonary tuberculosis. After a period of study under Karl Pearson, he was in 1910 (the year in which his first book, on the *Physiology of the Special Senses*, appeared) appointed statistician to the Lister Institute. Here he carried out statistical investigations into such diverse topics as the fatality of fractures and pneumonia in hospital practice, the epidemiology of plague, factors influencing rates of infant mortality, tuberculosis and cancer and, an early indication of his ultimate chief field of effort, an examination of factors determining the rise and spread of epidemic diseases. In the first World War he served for a time with the rank of Captain in the R.A.M.C. until 1917, when he was put in charge of the medical research subsection of the Ministry of Munitions. Here, naturally, his statistical energies were switched to the special problems of the hour, such as dietaries and energy expenditure in relation to munition workers, the incidence of tuberculosis in industry, the factors influencing labour wastage and sickness and accident frequency, but in the more general field he contributed to the study of problems associated with cerebro-spinal fever and influenza. In this period began his associations with Yule, Collis and May Smith on the industrial problems of accidents, sickness and fatigue. His book, with Collis, on the *Health of the Industrial Worker*, appeared in 1921, and in 1935, in the Heath Clark lectures, he returned to the discussion of the influence of environmental factors in industry.

In 1919 he was put in charge of the medical statistics section of the newly created Ministry of Health, a post which he retained until 1928, the section being accommodated in the National Institute for Medical Research at Hampstead. Here, apart from the routine epidemiological and other statistical work for the Ministry, much of which appeared in unsigned contributions to official reports, the diversity and intensity of his labours are

reflected in over sixty individual or collaborate papers. In the earlier half of these ten years, in collaboration first with Woods and later with Yule, appeared his first writings on the frequency of multiple accidents and disease, work which led to so much fruitful subsequent investigation by others and which ranks among his most important contributions to statistical medicine. At this time, too, his deep interest in the historical aspects of epidemiology became apparent in his dissertations on the contributions made in this field by Sydenham and Galen. Among others of his researches at this period must be mentioned those on malignant disease, especially on the statistical aspects of the evaluation of different methods of therapy. Apart from their direct contributions to knowledge of the subject, these revealed the inadequacies of existing, and the potentialities of properly kept hospital records in throwing light on medical problems especially of disease description and therapeutic assessment. In the wider field of routine medical statistics he produced at this time, with Granville Edge and on behalf of the League of Nations Health Organization, a series of booklets surveying the nature and content of the official vital statistics available in many European countries, designed to guide the unwary in the treacherous field of international comparisons. In the later part of his stay at the National Institute appeared the first of the series of joint papers with Topley on experimental epidemiology. The difficulties of human epidemiology might, it was thought, be lessened in study of populations of mice under strict control ('a little epidemiological world of our own nearer to the heart's desire'), the language of epidemiology might be more readily learned and the findings utilized in the interpretation of the more complicated happenings in man. Although fifteen years of constant questioning failed to produce all the answers, it did sensibly clarify many aspects of epidemic occurrences and revealed much of practical value, for example, on the limitation of protection by inoculation. This, perhaps the main single work of his life, was continued after his appointment in 1928 to the chair of Vital Statistics and Epidemiology at the London School of Hygiene and Tropical Medicine, where he remained until his retirement in 1945. The work is described in his Herter lectures in 1931, in *Epidemics and Crowd Diseases* (1935), in *Experimental Epidemiology*, M.R.C. Report No. 209 (1936) and in a 'Statistical study of infectious diseases' in the *Journal of the Royal Statistical Society* (1946). The switching of interest in recent years to the new therapeutic and prophylactic procedures in infectious diseases, and the pronounced success attendant on these new methods in reducing incidence and mortality has temporarily overshadowed the importance of this brilliant and painstaking teamwork. Following the death of Brownlee, Greenwood undertook the direction of the statistical staff of the M.R.C. and, on its formation, became chairman of its statistical committee, remaining so until 1948. The work of this committee, with members representative of the different kinds of statistical acumen, exercised considerable influence on medical research, and was a powerful stimulus to recognition of the need for and the more rapid diffusion of statistical knowledge throughout the profession. At the School of Hygiene and Tropical Medicine, in addition to the continuation of his long-term work with Topley, his interests widened and, both by his teaching and writings, his influence in and on the profession steadily increased. It is not surprising that many honours were bestowed on him. He was awarded the Buchanan Medal of the Royal Society in 1927 and the Guy Medal in gold of the Royal Statistical Society in 1945; he was Milroy lecturer (Royal College of Physicians) in 1922 and Herter lecturer in Baltimore in 1931. In 1924 he was elected to the Fellowship of the Royal College of Physicians and of the Royal Society in 1928. From 1919 to 1934 he was honorary secretary of the Royal Statistical

Society and became its president in 1934–6. He was an honorary member of the American Statistical Association and of the Indian National Academy of Science, and a member of the International Statistical Institute. In later years Greenwood was associated with *Biometrika*, first in an advisory editorial capacity and from 1946 until his death as one of the Trustees.

It is difficult at present to assess his contributions to the advance of knowledge—as one of the earliest medical men adequately equipped statistically, his investigations, individual and collaborate, perhaps naturally covered a multitude of individual subjects. Even from his earliest years, however, his principal interests were in the problems of epidemic disease, and probably his contributions here, together with his studies on industrial sickness and accidents, will remain his most abiding claims. Yet more important than his writings, perhaps, is the success with which he laboured throughout his professional life to introduce statistical ideas and methods into medical investigation. The opportunities afforded for this by his associations with the research and teaching organizations in London were eagerly grasped by him, and the many who sought his advice must feel indebted for the inspiration and guidance so consistently obtained. In those fortunate enough to have been more intimately associated with him remain the deep affection and respect offered only to the great.

P. L. MCKINLAY

TABLES OF THE 5% AND 0·5% POINTS OF PEARSON CURVES (WITH ARGUMENT β_1 AND β_2) EXPRESSED IN STANDARD MEASURE

By E. S. PEARSON and MAXINE MERRINGTON

1. The purpose and form of the tables

A considerable proportion of current statistical research is concerned with the determination of the sampling distributions of statistics required either as estimators or for use in tests of significance. It is often the case that while the distribution itself cannot be expressed in any simple form, the sampling moments can be derived and numerical values calculated either precisely or as approximations. In a great number of cases such investigation shows that the distribution tends to the normal form as the sample size is increased. However, this information regarding the limit is often not sufficient, and we require an answer to the question: what error is involved in assuming normality when the sample size has a specified value, n? In other words, to what extent in practice will the knowledge of the expectation and standard error of a particular statistic suffice?

Accumulated experience has shown that the moment ratios

$$\sqrt{\beta_1} = \gamma_1 = \mu_3/\mu_2^{\frac{3}{2}}, \quad \beta_2 = \gamma_2 + 3 = \mu_4/\mu_2^2 \tag{1}$$

provide extremely useful indices of departure from normality, but it is clear that the knowledge that a certain statistic T, has say, for $n = 20$, a sampling distribution with $\sqrt{\beta_1} = 0{\cdot}09$, $\beta_2 = 3{\cdot}21$, does not immediately help us in assessing the error involved in assuming that T is normally distributed. There can be no complete answer to this question since the first four moments do not in general specify a frequency function. However, we have thought it useful to make available tables which will relate certain percentage points, expressed in standard measure, to the β_1 and β_2 of the distribution, for curves of Karl Pearson's system. These curves, derived from the differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-(x+c_1)}{c_0 + c_1 x + c_2 x^2}, \tag{2}$$

while differing in algebraic form, are continuous in shape for changes in β_1 and β_2. If the frequency curve is denoted by $y = f(x), \quad b_1 \leqslant x \leqslant b_2,$

and x has expectation μ and standard deviation σ, then we have tabulated the values of $u = (x_\alpha - \mu)/\sigma$, where

$$\int_{b_1}^{x_\alpha} f(x)\,dx = \alpha, \tag{3}$$

for $\alpha = 0{\cdot}005$, $0{\cdot}05$, $0{\cdot}95$ and $0{\cdot}995$. This we have done for the combination of values of β_1 and β_2 shown in the tables below. We have used $\beta_1 = \gamma_1^2$ rather than $\sqrt{\beta_1}$ as argument in order to make use of an earlier tabulation (Pearson, 1925, pp. 439–42, and *Tables for Statisticians and Biometricians*, vol. **2**, Tables L); the skewness has been assumed positive, i.e. $\mu_3 \geqslant 0$.

We should have liked to have included tables giving additional percentage points, e.g. for α and $1-\alpha$ equalling $0{\cdot}01$ and $0{\cdot}025$, and it is hoped to compute these at a later date. In the first instance, however, the choice of the levels, $0{\cdot}005$ and $0{\cdot}05$, was fixed for us by the earlier work. With these two points known at each end of the distribution, the broad effect of departure from normality can be judged.

Our reasons for believing that the percentage points of Pearson curves, tabulated according to β_1 and β_2, will provide useful approximations to the corresponding points of other distributions for which only the sampling moments are known, may be summarized as follows:

(*a*) These curves represent the exact distribution of a number of the more important statistics in common use, a property which is not shared, for instance, by curves of the Gram-Charlier system. This suggests a general appropriateness of form.

(*b*) In several cases percentage points have first been calculated using a Pearson curve with the correct first four moments, and later computational work has substantially confirmed the accuracy of the approximation. This occurred in the case of the distribution of range, where the table of percentage points based on Type I and VI curves in use from 1932 was superseded for $n \leqslant 12$ by the exact calculations of 1942 (Pearson, 1932; Hartley & Pearson, 1942). In other cases, for example with forms of the likelihood ratio developed by Wilks, the adequacy of approximations based on this system of curves has been checked by comparison in certain special cases where the true distribution is known.

It is in the Type IV area of the β_1, β_2-plane (see Fig. 1) where there is at present least evidence regarding the appropriateness of this curve system in sampling theory. For β's in the Type IV area there are a number of alternative curve systems, and work aimed at comparing their different properties is in progress. But whatever the outcome of such research, there can be little doubt that within this area of the β_1, β_2-plane, the Type IV approximation will give far more accurate values to the percentage points than would be obtained from the normal probability scale.

2. ILLUSTRATIONS OF THE USE OF THE TABLES

Example 1

Bailey (1950), in considering the mathematical theory of what he terms a simple stochastic epidemic, gives a table of moments for the time of completion of an epidemic, when all available susceptibles have been exhausted. Thus if τ is the completion time in a community containing n susceptibles, he shows that appreciable skewness and kurtosis remain in the distribution of τ, even when n is very large. With $n = 80$, he gives:

γ_1	γ_2	$\beta_1 = \gamma_1^2$	$\beta_2 = \gamma_2 + 3$
0·771	1·114	0·594	4·114

The following figures for percentage levels of $(\tau - \bar{\tau})/\sigma_\tau$ may be derived by interpolation in Tables 1–4 below:

	Lower		Upper	
	0·5%	5%	5%	0·5%
Using Type IV approx.	−1·95	−1·41	1·82	3·33
From 'normal' probability scale	−2·58	−1·64	1·64	2·58

The figures give added meaning to Bailey's measures of skewness and kurtosis and a comparison with the normal-scale percentage levels is instructive.

Table 1. *Lower 5 % points of the standardized deviate* $(x_\alpha - \mu)/\sigma$, $(\alpha = 0{\cdot}05)$

(Note that for positive skewness, i.e. $\mu_3 > 0$, the deviates in this table are negative.)

β_2 \ β_1	0·00	0·01	0·03	0·05	0·10	0·15	0·20	0·30	0·40	0·50	0·60	0·70	0·80	0·90	1·00
1·8	1·56	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2·0	1·61	1·56	1·51	1·47	—	—	—	—	—	—	—	—	—	—	—
2·2	1·64	1·59	1·55	1·52	1·46	1·40	1·35	—	—	—	—	—	—	—	—
2·4	1·65	1·61	1·58	1·55	1·50	1·45	1·41	1·33	—	—	—	—	—	—	—
2·6	1·65	1·61	1·59	1·57	1·53	1·49	1·45	1·38	1·30	—	—	—	—	—	—
2·8	1·65	1·62	1·59	1·57	1·54	1·51	1·48	1·42	1·35	1·29	—	—	—	—	—
3·0	1·64	1·62	1·59	1·58	1·55	1·52	1·49	1·44	1·39	1·33	1·27	—	—	—	—
3·2	1·64	1·61	1·59	1·58	1·55	1·53	1·50	1·46	1·42	1·37	1·31	1·25	1·19	—	—
3·4	1·64	1·61	1·59	1·58	1·55	1·53	1·51	1·47	1·43	1·39	1·35	1·30	1·24	1·19	—
3·6	1·63	1·61	1·59	1·58	1·55	1·53	1·52	1·48	1·45	1·41	1·37	1·33	1·28	1·23	1·18
3·8	1·63	1·60	1·59	1·58	1·55	1·54	1·52	1·49	1·46	1·42	1·39	1·35	1·31	1·27	1·23
4·0	1·62	1·60	1·59	1·57	1·55	1·54	1·52	1·49	1·46	1·43	1·40	1·37	1·34	1·30	1·26
4·2	1·62	1·60	1·58	1·57	1·55	1·54	1·52	1·49	1·47	1·44	1·42	1·39	1·36	1·32	1·28
4·4	1·62	1·60	1·58	1·57	1·55	1·54	1·52	1·50	1·47	1·45	1·42	1·40	1·37	1·34	1·31
4·6	1·61	1·59	1·58	1·57	1·55	1·54	1·52	1·50	1·48	1·45	1·43	1·41	1·38	1·35	1·33
4·8	1·61	1·59	1·58	1·57	1·55	1·53	1·52	1·50	1·48	1·46	1·44	1·41	1·39	1·37	1·34
5·0	1·60	1·58	1·57	1·56	1·55	1·53	1·52	1·50	1·48	1·46	1·44	1·42	1·40	1·38	1·35

Table 2. *Upper 5 % points of the standardized deviate* $(x_\alpha - \mu)/\sigma$, $(\alpha = 0{\cdot}95)$

β_2 \ β_1	0·00	0·01	0·03	0·05	0·10	0·15	0·20	0·30	0·40	0·50	0·60	0·70	0·80	0·90	1·00
1·8	1·56	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2·0	1·61	1·66	1·70	1·72	—	—	—	—	—	—	—	—	—	—	—
2·2	1·64	1·68	1·71	1·74	1·77	1·80	1·83	—	—	—	—	—	—	—	—
2·4	1·65	1·69	1·71	1·74	1·77	1·80	1·83	1·87	—	—	—	—	—	—	—
2·6	1·65	1·68	1·71	1·73	1·76	1·79	1·81	1·86	1·90	—	—	—	—	—	—
2·8	1·65	1·68	1·70	1·72	1·75	1·77	1·80	1·84	1·88	1·92	—	—	—	—	—
3·0	1·64	1·67	1·69	1·71	1·74	1·76	1·78	1·82	1·86	1·90	1·93	—	—	—	—
3·2	1·64	1·67	1·69	1·70	1·73	1·75	1·77	1·80	1·84	1·87	1·91	1·94	1·98	—	—
3·4	1·64	1·66	1·68	1·69	1·72	1·74	1·76	1·79	1·82	1·85	1·88	1·92	1·95	1·98	—
3·6	1·63	1·65	1·67	1·68	1·71	1·73	1·74	1·77	1·80	1·83	1·86	1·89	1·92	1·95	1·98
3·8	1·63	1·65	1·66	1·68	1·70	1·72	1·73	1·76	1·79	1·82	1·84	1·87	1·90	1·93	1·96
4·0	1·62	1·64	1·66	1·67	1·69	1·71	1·72	1·75	1·78	1·80	1·83	1·85	1·88	1·90	1·93
4·2	1·62	1·64	1·65	1·66	1·68	1·70	1·71	1·74	1·76	1·79	1·81	1·83	1·86	1·88	1·91
4·4	1·62	1·63	1·65	1·66	1·68	1·69	1·70	1·73	1·75	1·78	1·80	1·82	1·84	1·86	1·89
4·6	1·61	1·63	1·64	1·65	1·67	1·68	1·70	1·72	1·74	1·76	1·78	1·80	1·82	1·84	1·87
4·8	1·61	1·62	1·64	1·65	1·66	1·68	1·69	1·71	1·73	1·75	1·77	1·79	1·81	1·83	1·85
5·0	1·60	1·62	1·63	1·64	1·66	1·67	1·68	1·71	1·73	1·74	1·76	1·78	1·80	1·82	1·84

Table 3. *Lower* 0·5 % *points of the standardized deviate* $(x_\alpha - \mu)/\sigma$, $(\alpha = 0{\cdot}005)$

(Note that for positive skewness, i.e. $\mu_3 > 0$, the deviates in this table are negative.)

β_2 \ β_1	0·00	0·01	0·03	0·05	0·10	0·15	0·20	0·30	0·40	0·50	0·60	0·70	0·80	0·90	1·00
1·8	1·71	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2·0	1·92	1·80	1·71	1·64	—	—	—	—	—	—	—	—	—	—	—
2·2	2·10	1·99	1·89	1·82	1·68	1·56	1·46	—	—	—	—	—	—	—	—
2·4	2·26	2·14	2·04	1·97	1·83	1·71	1·62	1·44	—	—	—	—	—	—	—
2·6	2·38	2·27	2·18	2·12	1·98	1·87	1·77	1·58	1·42	—	—	—	—	—	—
2·8	2·49	2·38	2·30	2·23	2·10	1·99	1·89	1·71	1·55	1·41	—	—	—	—	—
3·0	2·58	2·48	2·39	2·33	2·21	2·11	2·01	1·84	1·68	1·53	1·40	—	—	—	—
3·2	2·65	2·55	2·48	2·42	2·30	2·20	2·11	1·95	1·79	1·65	1·51	1·39	1·27	—	—
3·4	2·71	2·61	2·54	2·48	2·38	2·28	2·20	2·04	1·90	1·76	1·62	1·50	1·38	1·27	—
3·6	2·76	2·67	2·60	2·54	2·44	2·35	2·27	2·13	1·99	1·85	1·72	1·60	1·48	1·37	1·27
3·8	2·80	2·71	2·65	2·60	2·50	2·41	2·34	2·20	2·07	1·94	1·82	1·70	1·58	1·47	1·37
4·0	2·83	2·75	2·69	2·64	2·54	2·47	2·39	2·26	2·14	2·02	1·90	1·78	1·67	1·56	1·45
4·2	2·87	2·79	2·72	2·68	2·59	2·51	2·44	2·32	2·20	2·09	1·97	1·86	1·75	1·65	1·54
4·4	2·90	2·82	2·76	2·71	2·62	2·55	2·49	2·37	2·25	2·15	2·04	1·93	1·83	1·73	1·62
4·6	2·92	2·85	2·79	2·74	2·66	2·59	2·52	2·41	2·30	2·20	2·10	2·00	1·90	1·80	1·70
4·8	2·94	2·87	2·81	2·77	2·69	2·62	2·56	2·45	2·35	2·25	2·15	2·05	1·96	1·87	1·77
5·0	2·96	2·89	2·83	2·79	2·71	2·65	2·59	2·48	2·39	2·29	2·20	2·11	2·01	1·92	1·84

Table 4. *Upper* 0·5 % *points of the standardized deviate* $(x_\alpha - \mu)/\sigma$, $(\alpha = 0{\cdot}995)$

β_2 \ β_1	0·00	0·01	0·03	0·05	0·10	0·15	0·20	0·30	0·40	0·50	0·60	0·70	0·80	0·90	1·00
1·8	1·71	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2·0	1·92	2·01	2·06	2·09	—	—	—	—	—	—	—	—	—	—	—
2·2	2·10	2·19	2·24	2·27	2·31	2·33	2·35	—	—	—	—	—	—	—	—
2·4	2·26	2·35	2·41	2·44	2·49	2·52	2·53	2·53	—	—	—	—	—	—	—
2·6	2·38	2·48	2·54	2·57	2·63	2·66	2·68	2·70	2·69	—	—	—	—	—	—
2·8	2·49	2·58	2·64	2·68	2·73	2·77	2·80	2·83	2·84	2·83	—	—	—	—	—
3·0	2·58	2·66	2·72	2·76	2·82	2·86	2·89	2·93	2·95	2·96	2·95	—	—	—	—
3·2	2·65	2·73	2·79	2·83	2·89	2·93	2·96	3·01	3·04	3·06	3·07	3·06	3·04	—	—
3·4	2·71	2·79	2·85	2·88	2·95	2·99	3·02	3·07	3·11	3·13	3·15	3·16	3·15	3·14	—
3·6	2·76	2·84	2·89	2·93	2·99	3·03	3·07	3·12	3·16	3·19	3·22	3·23	3·24	3·24	3·23
3·8	2·80	2·88	2·93	2·97	3·03	3·07	3·11	3·16	3·20	3·24	3·27	3·29	3·30	3·31	3·32
4·0	2·83	2·91	2·96	3·00	3·06	3·10	3·14	3·20	3·24	3·28	3·31	3·34	3·36	3·37	3·38
4·2	2·87	2·94	2·99	3·03	3·09	3·13	3·17	3·22	3·27	3·31	3·34	3·37	3·40	3·42	3·43
4·4	2·90	2·97	3·02	3·05	3·11	3·15	3·19	3·25	3·29	3·33	3·37	3·40	3·42	3·45	3·47
4·6	2·92	2·99	3·04	3·07	3·13	3·17	3·21	3·27	3·31	3·36	3·39	3·42	3·44	3·47	3·50
4·8	2·94	3·01	3·06	3·09	3·15	3·19	3·23	3·28	3·33	3·37	3·41	3·44	3·47	3·49	3·52
5·0	2·96	3·03	3·07	3·11	3·16	3·21	3·24	3·30	3·35	3·39	3·43	3·46	3·49	3·52	3·54

Example 2

If we write $$m = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$$

for the mean deviation in a random sample of observations $x_1, x_2, \ldots, x_n$ drawn from a normal population, exact expressions for the first two sampling moments of m are known. Using these formulae for $\overline{m}$ and σ_m and some tabulated values for β_1 and β_2 derived from expansions given by Geary (see Editorial Notes, *Biometrika*, 1946, 1948), we have the following values for the cases $n = 10$ and 20:

n	$\overline{m}$	σ_m	$\beta_1(m)$	$\beta_2(m)$
10	0·75694	0·18945	0·1064	3·093
20	0·77768	0·13438	0·0513	3·045

We may now determine approximations to the lower and upper 0·5 % points in these two sampling distributions, expressed in terms of the population standard deviation as unit. These will be given by $\overline{m} + f_\alpha \sigma_m$, where the factors f_α for $\alpha = 0{\cdot}005$ and 0·995 are obtained by interpolation from Tables 3 and 4 below. The results are as follows:

n	$f_{0\cdot005}$	$f_{0\cdot995}$	Lower 0·5 % point	Upper 0·5 % point
10	−2·24	2·86	0·333	1·299
20	−2·35	2·78	0·462	1·151

The values for $n = 10$ correspond, to three-decimal accuracy, with the exact percentage points derived from the probability integral of the mean deviation, as developed by Godwin (1946) and computed by Hartley (1946). The probability integral was not calculated beyond $n = 15$, but the accuracy of the approximation at $n = 10$ suggests that the values given at $n = 20$ should also be satisfactory.

3. Method of computation

The construction of the tables is most easily explained by reference to the accompanying Fig. 1. It was decided to limit the range of calculation to distributions having $\beta_1 \leqslant 1{\cdot}0$ and $\beta_2 \leqslant 5{\cdot}0$ ($\gamma_2 \leqslant 2{\cdot}0$).* Further, J and U curves were excluded so that the upper boundary of the region (in the sense of the chart) is the Type IX line corresponding to curves with equation

$$y = f(x) = y_0\left(1 + \frac{x}{a}\right)^m. \tag{4}$$

* The original calculations were made in connexion with a Type I approximation to the sum of the terms of a hypergeometric distribution. The percentage limits were given for Type I curves having β_1 as large as 3·0.

The tables previously referred to (Pearson, 1925, pp. 439–42 and *Tables for Statisticians and Biometricians*, vol. **2**, Tables L*a–d*) gave the standardized percentage limits for Type I curves lying in the area between the broken line shown in the chart and the Type III line, $2\beta_2 - 3\beta_1 - 6 = 0$. Considerable additional calculations were therefore needed.

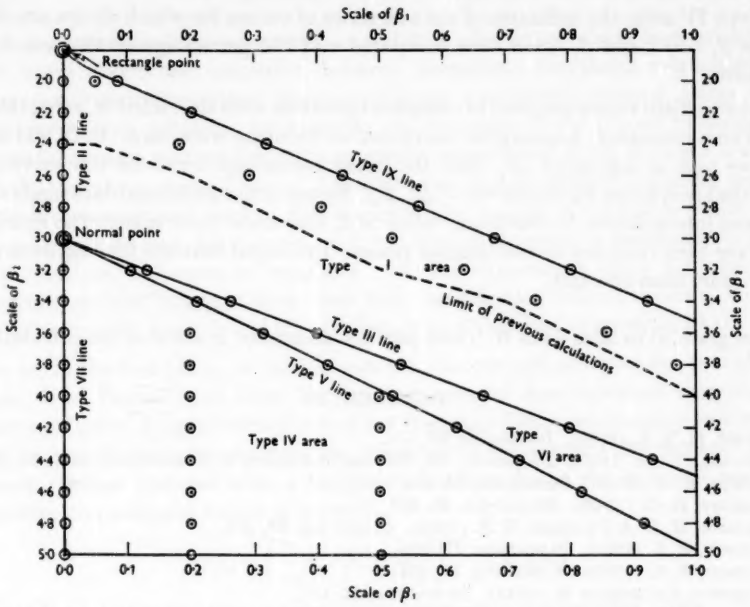


Fig. 1. Showing the area of the β_1, β_2 plane covered by the tables and the framework of curves used in their calculation.

Details of the earlier calculations have been described by one of us (Pearson, 1925, p. 438). To extend this work, a framework of β_1, β_2 points was used, as shown by small circles in Fig. 1. At each point the four percentage levels for the appropriate Pearson curve were computed to four-decimal accuracy. It will be seen that the points selected fall along certain lines; these lines and the method of procedure are indicated below:

(*a*) Symmetrical distributions, Types II and VII; by interpolation in tables of the percentage points of Student's *t*-distribution, using the appropriate fractional degrees of freedom.

(*b*) Type IX line; by integration of equation (4) above.

(*c*) A line of points shown below and nearly parallel to the Type IX line. The corresponding distributions are those with equations

$$f(x) = x^{0\cdot5}(1-x)^{q-1}/B(1\cdot5, q), \tag{5}$$

values of q being chosen to make $\beta_2 = 2\cdot4(0\cdot2)3\cdot8$. The lower percentage levels were then found from the tables of the percentage points of the Incomplete Beta Function (Catherine M. Thompson, 1941), by interpolation for $\nu_1 = 2q$ in the row with $\nu_2 = 2p = 3$. For the upper levels the roles of ν_1 and ν_2 were reversed, and $1-x$ read for x.

(*d*) The levels corresponding to $\beta_1 = 0, \beta_2 = 1{\cdot}8$ were obtained from a rectangle and those for $\beta_1 = 0{\cdot}05, \beta_2 = 2{\cdot}0$ by special quadrature.

(*e*) Type III line; by interpolation in tables of percentage points of the χ^2 distribution.

(*f*) Type V line; by use of the reciprocal transformation, which converts a Type V into a Type III variable, and so leads again to the use of the χ^2 tables.

(*g*) Type IV area; the ordinates of the two series of curves for which circles are shown on the lines $\beta_1 = 0{\cdot}2$ and $\beta_1 = 0{\cdot}5$ were computed and the percentage levels were found by quadrature.

All intermediate values required to complete the tables were then filled in, using this framework. Where necessary, Lagrangian interpolation formulae were used. It should be noted that if we take as argument $\sqrt{\beta_1}$, then the upper percentage levels for the curve $(\sqrt{\beta_1}, \beta_2)$ become the lower levels for the curve $(-\sqrt{\beta_1}, \beta_2)$. By use of the additional data made available in this way interpolation for the lower values of β_1 was made much easier. The figures in the tables have been recorded to two decimal places; it is hoped that the final figure is never in error by more than one unit.

We are grateful to Miss Jean H. Thompson for assistance in some of the calculations.

REFERENCES

Bailey, N. T. J. (1950). *Biometrika*, **37**, 193.
Editorial Note (1946). *Biometrika*, **33**, 252; and correction in *Biometrika* (1948), **35**, 424.
Godwin, H. J. (1946). *Biometrika*, **33**, 254.
Hartley, H. O. (1946). *Biometrika*, **33**, 257.
Hartley, H. O. & Pearson, E. S. (1942). *Biometrika*, **32**, 301.
Pearson, E. S. (1925). *Biometrika*, **17**, 388.
Pearson, E. S. (1932). *Biometrika*, **24**, 404.
Thompson, Catherine M. (1941). *Biometrika*, **32**, 151.

REGRESSION, STRUCTURE AND FUNCTIONAL RELATIONSHIP. PART I

By M. G. KENDALL

1. In this paper I shall try to remove a number of obscurities which have arisen in the past few years about the connexion between regression, functional relationship and structural equations in theoretical models containing a stochastic element. Some of these obscurities arise from notation; for example, we sometimes write $y = ax$ to denote both a purely mathematical relationship and a regression relationship, or we write $y = x + \epsilon$ while thinking of y and x as mathematical variables and ϵ as a random variate, and either procedure can give rise to nonsensical results. But there are also sources of confusion which are not purely notational and lead, for instance, to differences of approach in econometrics through 'errors in variables' or 'errors in equations'. Related expressions which give rise to misunderstanding are the 'fitting of lines when both variables are subject to error', the 'fixed variable' of regression analysis and the 'predetermined variable' of econometrics.

2. Let us, in the first place, go back to consider the concept of regression as understood by Galton, Karl Pearson and Yule. To them a regression line expressed a property of a bivariate population. It summarized one of the features of joint relationship in very much the same way as the mean summarizes the properties of a univariate distribution. If x and y represent random variates with a frequency function $f(x, y)$ the series of parameters corresponding to particular values of x, say X,

$$\mu'_{Xr} = \int_{-\infty}^{\infty} y^r f(X, y)\, dy \Big/ \int_{-\infty}^{\infty} f(X, y)\, dy \tag{1}$$

bears an obvious relation to the moments in the univariate case and the curve

$$Y_r = \mu'_{Xr} = \int_{-\infty}^{\infty} y^r f(X, y)\, dy \Big/ \int_{-\infty}^{\infty} f(X, y)\, dy \tag{2}$$

expresses the way in which the rth moment of y varies according to assigned values of x. In the particular case when $r = 1$ we have the so-called regression line of y on x

$$Y = \int_{-\infty}^{\infty} y f(X, y)\, dy \Big/ \int_{-\infty}^{\infty} f(X, y)\, dy. \tag{3}$$

I shall throughout use capitals such as X, Y to denote only mathematical variables such as current co-ordinates and x, y to denote random variates; and I distinguish between the *variable*, which means a mathematical variable in the ordinary sense, and a *variate*, which is associated with a frequency or probability distribution of possible values.

3. The regression line (Y, X), from this point of view, is in no sense a functional relationship between y and x. We may regard it as a functional relationship between Y and X; given any value X the integral on the right in equation (3) determines a value of Y. But taken in conjunction with the underlying frequency distribution, equation (3) is more than a functional relation; for we are to regard some values of x as occurring with greater probability than others.

4. The point becomes clear when we examine a statement which is often made in textbooks of statistics (including, I regret to say, one of my own) that when the correlation

coefficient is unity there is a functional relation between the variables. This is true only in a rather special sense. There is merely a functional relation between the variates. If we have two (mathematical) variables Y and X defined in a certain range, it is statistical nonsense to speak of their distribution and still more so to speak of their variance and covariance or of their correlation. A variable has no frequency function associated with it. We can, of course, associate such a function and sometimes we do so; but in the act of doing so we turn the (mathematical) variables Y and X into (statistical) variates x and y. If the product-moment correlation of these variates x and y is unity there is a linear relation between x and y. One statistical variate, measured about its mean, is then a constant times the other; but there is still an underlying frequency function common to both. To put it geometrically, if a bivariate frequency surface shrinks or collapses into one dimension there remains a univariate frequency curve which expresses the relative frequency or probability with which values of the variate arise. All that has happened is that two-way variation has contracted into one-way variation on a scale which is functionally connected with the two original scale variables X and Y. It will be seen later, I hope, that I am not splitting hairs.

5. I repeat then, that to the founders of regression theory the regression curves were summarizing aspects of an underlying multivariate distribution. This outlook dominated orthodox statistical thought until the early 1920's. At that stage we may detect several lines of departure; or perhaps it would be better to say, several extraneous lines of development converging on to the topic of regression. The fitting of curves to series of points when both variables are subject to error; the development of component- and factor-analysis; the type of problem which led Frisch to introduce confluence analysis; these and other subjects aim in common at the production of functional relationships, in some sense or other, between a complex of variables or variates, and they have all become entangled more or less with the regression concept. The change in outlook on regression analysis itself was subtle and is difficult to trace in the literature because it never (so far as I know) found explicit expression. There was, however, such a change. It may perhaps be brought out most clearly by an examination of what is meant by a 'fixed variable', or, as I should prefer to say, a 'fixed' variate.

6. We suppose ourselves given a bivariate population specified according to two variates x and y. For non-statistical reasons we regard one as 'dependent', the other as an 'independent' variate. These terms are themselves misleading and tendentious from the older point of view, a fact which has led to the introduction of other words with a similar (though not, apparently, an identical) connotation such as 'explanatory' or 'predetermined' variable.* I do not like adding to the terminology of an overburdened subject, but to be absolutely clear I shall, at least for the purposes of this article, use a new word. The variate x which enters into the right-hand side of equations of type (1) will be called the *predicated* variate. The variate y, which also enters into the right-hand side but is integrated out, will be called the *unpredicated* variate. For multiple variation we may have more than one predicated variate, but there will never be more than one unpredicated variate.† The corresponding variables Y and X may also be said to be unpredicated or predicated respectively.

* In the theory of stochastic processes, when a sequence of random functions depends on a variable t, this is also referred to sometimes as the 'independent variable'.

† Regression theory may be extended to the regression of vectors on vectors, in which case it might be said that we have several unpredicated scalars making up the unpredicated vector.

With this convention, we may say that the object of regression analysis for two variates y and x is to express the mean of the unpredicated variate y as a function of the predicated variable X; or, reciprocally, the mean of x as a function of Y.

7. In the classical approach questions of sampling usually arose in circumstances where the sampled members were chosen by a method which did not depend on the values of the variates they bore. There was no question of errors of measurement. The sample was drawn or given, the values of the variates determined for each member, one variate selected as the predicated variate for non-statistical reasons, and the regression line then estimated. In more recent developments, however, it has been usual to regard the values of the predicated variate as determinate before the sample is drawn; that is to say, we fix certain values of X and observe the corresponding values of y. Instead of taking members at random with respect to both variates we assign values of one and sample at random with respect to the other.

8. It is easy to imagine circumstances where this procedure has decisive advantages and even where it is practically necessary and, conversely, where it has decisive disadvantages or is impossible.

(*a*) If we know that the regression is linear and have only a few observations to make, it may be better to arrange for them to be made near the two tails of the x-distribution instead of near the mean, which would give less accuracy to the sample regression line.

(*b*) We may be compelled to choose members according to the values of one of the variates under consideration. In an inquiry into the relation between personal income and personal savings, for example, we may require to stratify by the income variate. In considering the relation between performance in final examinations and performance at entrance examinations in a group of students we often have only a truncated part of the predicated variate (entrance performance) because students who fail the entrance do not proceed to the final examination.

(*c*) On the other hand, if pre-selection of values of the predicated variate destroys or seriously impairs the validity of significance tests we may purchase extra accuracy too dearly. We then arrive at the familiar dilemma in the systematization of design—whether we prefer to be closer to the truth without knowing how close, or to run the risk of being further from it while knowing how far away we are likely to be.

(*d*) It should be observed that I do not adduce as an example the case occurring frequently in physics where we assign values of one variable and measure the values of the other; as, for example, in a laboratory experiment on Hooke's law, where we hang pre-selected weights on a spring and measure the extension. This looks like a case where the weight variable is pre-selected completely. But, in fact, it does not fall within the ambit of regression analysis at all unless there are no errors in the determination of the weights.

9. Let us suppose for simplicity that the regression of y on x is linear and that the variates are measured about their mean. The regression line is then

$$Y = \beta_0 + \beta_1 X, \tag{4}$$

where β_0 and β_1 are unknown constants. We can also write

$$y_X = \beta_0 + \beta_1 X + \epsilon_X, \tag{5}$$

meaning that the random variate y_X depends on a mathematical variable X, two unknown constants β_0 and β_1 and a random variate ϵ which also depends on X. What we should not write for the regression line is an expression such as

$$y = \beta_0 + \beta_1 x + \epsilon. \tag{6}$$

This means something quite different, namely, that y is the weighted sum of two random variates x and ϵ. It also seems to me confusing in this context, if not actually improper, to introduce the 'conditional' variate $y|x$ and write

$$y|x = \beta_0 + \beta_1 x + \epsilon|x, \tag{7}$$

which means that x is a random variate, that for any value of x, $y|x$ and $\epsilon|x$ are random variates, and that $y|x$ is a weighted sum of x and $\epsilon|x$.

The difference between (5) and (7) is that (5) asserts something about a random variate y and a variable X, whereas (7) asserts something about a pair of random variates y and x.

10. Consider now the problem of estimating the constants β_0 and β_1 in (4) from a random sample of values of y and x. We have two cases to consider: (1) when the sample is drawn so that values of y and x both occur at random, (2) when it is drawn by specifying values of x and only y varies at random.

To make any headway at all we require certain further assumptions. We may be content to rely on the principle of least squares, without looking for any deeper justification of that principle, and hence to minimize the sum of squares

$$\sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2,$$

which leads to the familiar estimators

$$b_0 = 0, \quad b_1 = \frac{\Sigma xy}{\Sigma x^2}, \tag{8}$$

both variates being measured about their observed mean. There is something to be said for accepting the principle of least squares in its own right (cf. Lindley, 1947). If, however, we do not accept least squares as a basic principle of estimation we must make certain assumptions about the random variate ϵ_X. What we do in practice is to suppose (1) that ϵ_X is independent of X; (2) that it is a normal variable with zero mean and variance, say σ^2; (3) that the observed values of ϵ_X are independent of each other. The joint distribution of a set of n ϵ's, whatever the values of the X's, is then

$$\frac{1}{(2\pi\sigma)^{\frac{1}{2}n}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} \epsilon_i^2\right\} d\epsilon_1 \dots d\epsilon_n. \tag{9}$$

We now use (5) to transform this to the random variables y_X and obtain

$$\frac{1}{(2\pi\sigma)^{\frac{1}{2}n}} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(y_X - \beta_0 - \beta_1 X)^2\right\} dy_{X,1} \dots dy_{X,n}, \tag{10}$$

We then use the principle of maximum likelihood to determine β_0 and β_1 so as to minimize the sum inside the braces, which leads us back to (8). It does not matter if we are ignorant of the value of σ.

11. This process is legitimate only if we use equation (5) to transform from the random variable element $d\epsilon_1 \dots d\epsilon_n$ to the element $dy_{X,1} \dots dy_{X,n}$. It is not legitimate if we use (6) or (7), for we are not entitled to express an n-fold element $d\epsilon_1 \dots d\epsilon_n$ in terms of an element comprising $2n$ variates, n y's and n x's. The method therefore provides a solution only in the case when values of X are given. What can be said about the alternative case when y and x are both random variates?

The usual answer to this question is that 'we may regard the x's as fixed' by an argument of the following type. Suppose the joint distribution of y and x is $f(x, y)$. The probability of the observed sample is then

$$P = f(x_1, y_1) \dots f(x_n, y_n)\, dx_1 \dots dx_n dy_1 \dots dy_n. \tag{11}$$

Transform to new random variates by the equations

$$\left.\begin{aligned} x &= \xi, \\ y &= \beta_0 + \beta_1 \xi + \epsilon. \end{aligned}\right\} \tag{12}$$

The Jacobian of the transformation is

$$\frac{\partial(x, y)}{\partial(\xi, \epsilon)} = 1,$$

and hence $$P = f(\xi_1, \beta_0 + \beta_1 \xi_1 + \epsilon_1) \dots f(\xi_n, \beta_0 + \beta_1 \xi_n + \epsilon_n)\, d\xi_1 \dots d\xi_n d\epsilon_1 \dots d\epsilon_n. \tag{13}$$

If this splits into two factors such that the ϵ-distribution is independent of the ξ-distribution we must have

$$f(\xi, \beta_0 + \beta_1 \xi + \epsilon) = g(\xi)\, h(\epsilon), \tag{14}$$

where g and h are frequency functions. This is equivalent to

$$f(x, y) = g(x)\, h(y - \beta_0 - \beta_1 x). \tag{15}$$

If and only if this condition is satisfied we may estimate β_0 and β_1 from $h(\epsilon)$ alone. In particular, if ϵ is distributed normally the maximum likelihood solution leads back to (9) and thence to (8). Thus, as the random origin of the x's does not affect the estimation we can regard them as the assigned values of a variable or as 'fixed'.

12. I cannot see any other way of arriving at the desired result unless we import yet another principle of inference, which is usually a confession of failure and in any case is not to be done where avoidable. The points to be noticed are: (1) that we have assumed the ϵ's to be independent of the x's and (2) that we have used in equations (12) a transformation depending on the view that both x and y are random variates. It is usually stated that 'by regarding the x's as fixed' we avoid making any assumption about the distribution of x; and, indeed, it is sometimes implied that the estimation remains valid whatever the distribution of x. This is true when the hypothesis as to linearity is satisfied, but not necessarily in alternative cases.

In questions of estimation we are mainly concerned with the distribution of the ϵ's. Where tests of significance are required we are interested in the distribution of the residuals about a *fitted* regression line, which is quite another thing.

13. Where the values of the predicated variate are decided before the sample is drawn I shall call the variate *determined*, in the contrary case *undetermined*. The latter is the classical case; the former is the type more frequently considered of recent years, particularly since Fisher's (1922) derivation of the t-test for regression coefficients.

It will be observed that in the determined case the problem of finding a line of type $Y = \beta_0 + \beta_1 X$ is not a bivariate problem at all. To make the contrast as sharp as possible we might even say that it is not a regression problem as understood by Galton and Pearson. It considers, not a pair of random variates, but a single random variate y_X depending on a mathematical* variable X. To arrive at the ordinary estimate (8) of the coefficients β_0 and β_1 it assumes that the residual ϵ_X is normal with constant variance. We also require the

* I should have used the customary word 'parameter' here to denote X but for the fact that 'parameter' has a special meaning in statistics. One sometimes feels that a Spirit of Perversity has been at work on statistical nomenclature in order to create difficulty and confusion.

assumption that under the system which generated the sample, values of y_X for the different value of X were chosen independently, or equivalently that ϵ_X for any particular value of X is independent of ϵ_X for other values.

14. In the undetermined case we require the condition expressed by equation (15). It follows that to arrive at the estimator (8) we cannot in every case regard the predicated variate as determined but that we may do so if (15) is satisfied. In that case it follows that the regression is linear. For any fixed X we have

$$\begin{aligned} E(y) &= \int_{-\infty}^{\infty} yh(y-\beta_0-\beta_1 X)\,dy \\ &= \int_{-\infty}^{\infty} (y-\beta_0-\beta_1 X)\,h(y-\beta_0-\beta_1 X)\,dy + (\xi_v+\beta_1 X)\int_{-\infty}^{\infty} h(y-\beta_0-\beta_1 X)\,dy \\ &= \beta_0+\beta_1 X, \end{aligned}$$

provided that the mean

$$\int_{-\infty}^{\infty} th(t)\,dt$$

is zero, which we may take to be so without loss of generality.

It is of some interest to examine the consequences of (15) more closely.

We have

$$\begin{aligned} \iint e^{it_1x+it_2y} f(x,y)\,dx\,dy &= \iint e^{it_1x+it_2y} g(x)\,h(y-\beta_0-\beta_1 x)\,dx\,dy \\ &= \iint e^{it_1x+it_2(\beta_0+\beta_1x)}\, e^{it_2(y-\beta_0-\beta_1x)}\, g(x)\,h(y-\beta_0-\beta_1 x)\,dx\,dy, \end{aligned}$$

or

$$\phi(t_1,t_2) = \phi_g(t_1+t_2\beta_1)\,\phi_h(t_2)\,e^{it_2\beta_0}, \tag{16}$$

where ϕ, ϕ_g and ϕ_h are the respective characteristic functions of f, g and h. Taking logarithms and assuming the existence of cumulants of all orders we have

$$\Sigma \frac{\kappa_{rs}(it_1)^r(it_2)^s}{r!\,s!} = it_2\beta_0 + \Sigma \frac{\kappa_r' i^r(t_1+t_2\beta_1)^r}{r!} + \Sigma \frac{\kappa_r''(it_2)^r}{r!}, \tag{17}$$

where κ, κ' and κ'' refer respectively to f, g and h. From the coefficient of t_1^r we have

$$\kappa_{r0} = \kappa_r' \quad (r>0), \tag{18}$$

and from the coefficient of $t_1^r t_2$ we have

$$\kappa_{01} = \beta_0+\beta_1\kappa_1'+\kappa_1'' \quad (r=0), \tag{19}$$

$$\kappa_{r1} = \beta_1\kappa_{r+1}' \quad (r\geqslant 1), \tag{20}$$

Thus we have

$$\beta_1 = \frac{\kappa_{11}}{\kappa_2'}, \tag{21}$$

$$\frac{\kappa_{r1}}{\kappa_{11}} = \frac{\kappa_{r+1}'}{\kappa_2'} \quad (r\geqslant 1), \tag{22}$$

Equation (22) is Wicksell's condition for linearity of regression (cf. Kendall, 1948, p. 335), which under certain conditions is necessary as well as sufficient.

We notice that from equation (21) the constant β_1 is given as the ratio of the covariance of y and x to the variance of x in the parent population. The estimator (8) is therefore consistent.

15. We may, then, regard the undetermined case as equivalent to the determined case only if the parent regression is linear. This is irrespective of any assumption about normality in the residuals. The conclusion is in accordance with common-sense requirements; for if

the regression is not linear, where we fix the values of X clearly makes a difference (apart from sampling effects) to the line we shall find.

16. In order to make further progress in the matter of estimation in the undetermined case we still require the assumption of normality in the residuals. All the information available about β_1 is contained in the function $h(\epsilon_1)\dots h(\epsilon_n)\,d\epsilon_1\dots d\epsilon_n$, and if we make the usual assumptions and use the method of maximum likelihood we return once again to equation (8). It may be noted that in this case $\kappa''_r = 0$ for $r > 2$ and we shall have from (17)

$$\kappa_{rs} = \beta_1^s \kappa'_{r+s} \quad (r \neq 0), \tag{23}$$

$$\kappa_{0s} = \beta_1^s \kappa'_s \quad (s \neq 2), \tag{24}$$

$$\kappa_{02} = \beta_1^2 \kappa'_2 + \kappa''_2. \tag{25}$$

17. So much for estimation. Somewhat similar considerations apply to the use of the t-test of significance of a regression coefficient with one predicated variable. In the determined case we derive the test on the basis of fixed values of X. In the undetermined case we may use the test in the same way if and only if the parent regression is linear and the residuals are normal without making any further assumption about the distribution of x. In such a case the test may be derived directly without passing the argument, so to speak, through the determined to the undetermined case; I have given the derivation in the particular case of joint bivariate normality (Kendall, 1948, p. 348).

18. In applying the t-test to a regression coefficient or to the difference of two regression coefficients we are thus (if the test is to have any power) postulating linearity of regression in the *alternative* hypotheses under test. Furthermore, we require joint independence of the residuals ϵ and the x's. Such conditions do not hold, for example, in autoregressive series, or, more generally, in any stochastic scheme containing lagged terms. Nor will they hold in time-variation where the successive values of x are auto-correlated owing to a time-effect, even though the correlation be a 'nonsense' one. For this reason the estimation and testing of significance of regression coefficients in time series differs from the ordinary case and special methods are necessary.

19. Let us consider the question from the point of view of 'conditional' inference. The argument here, as I understand it, is as follows: We draw an undetermined sample by a random process. This gives us certain values $x_1, \dots, x_n$. We decide, before drawing the sample, that whatever these values prove to be, we will make an inference in the restricted population of values (y, x) for which only these values $x_1, \dots, x_n$ appear. We do not fix the values of x in advance in the sense of the determined case; but we decide that we will regard them as fixed and make a conditional inference based on that assumption.

In general, this type of procedure has certain dangers and disadvantages, but I do not propose to discuss it in generality. For present purposes it is enough to note that the conditional distribution of t is the same in form whatever the values of x (on the usual assumption about linearity of regression and normality of residuals) and hence that we can make a series of inferences in the t-distribution for different sets of x's without having to confine ourselves to one particular set. In this context, I think, the 'conditional' approach is only another way of expressing the approach through the determined to the undetermined case. But it is not always so, and I repeat that equation (7) is not the same as equation (5). In equation (5) we select the values of X beforehand. In equation (7) we let chance select them and then confine our attention to the set so selected. The logic of the inference is different.

20. I interpolate here a comment to link up the 'determined' regression approach with factor analysis and the analysis of variance. In variance analysis it is customary to consider a variate y which may belong to one of i families as a systematic plus a residual component in some such form as

$$y_i = \beta_i + \epsilon_i. \tag{26}$$

This is similar in form to the regression equation

$$y_X = \beta_X + \epsilon_X,$$

but differs in that the families i need no longer be representable on a continuous (and orderable) scale and that we do not impose any relation of linearity on the constants β_i such as is involved in the regression approach. To make a significance test, however, we do require the same assumption about normality and independence in the residuals ϵ.

21. In factor analysis we postulate a scheme similar to (26), but the psychologist has, as a rule, not felt able to make the assumption about the normality, independence and equal variance of ϵ_i for all i. He regards ϵ_i as a specific term peculiar to the ith family. This involves him in the very serious trouble that his equations are then essentially indeterminate, a difficulty which has not been satisfactorily resolved (cf. Kendall, 1950). The statistician working in other subjects can hardly afford to throw stones at the factor analyst on this account; for the statistician avoids the difficulty only by 'assuming it away', namely, by postulating qualities in his residual terms ϵ. Sometimes he has prior evidence to support the reasonableness of such a postulation. But it is fair comment, I think, to say that a good many users of variance analysis and regression analysis make the assumption merely because they cannot otherwise obtain an answer and without examining whether it is the right or reasonable thing to do.

22. Consider now the case of one unpredicated and p predicated variates $x_1 \dots x_p$. We may write the linear regression as

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p. \tag{27}$$

Once again we may regard the case as completely determined when the X's are specified and write

$$y_{X_1,\dots,X_p} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon_{X_1,\dots,X_p}, \tag{28}$$

We may also have a completely undetermined case when no X is specified, and a partially determined case when some variates are determined and some are not.

To estimate the β's from a sample of n in the completely determined case we assume that $\epsilon_{X_1 \dots X_p}$ is independent of the X's, has zero mean and is normal with variance σ^2, and that values of ϵ are independent among themselves. This leads us, via the method of maximum likelihood, to the series of equations (where x_0 is a dummy variable equal to unity)

$$\Sigma y x_j = \Sigma \beta_0 x_j + \beta_1 \Sigma x_1 x_j + \dots + \beta_p \Sigma x_p x_j = 0 \quad (j = 0, \dots, p), \tag{29}$$

Let us consider the case when the predicated variates are completely undetermined.

23. In the manner of § 11 we transform by the equations

$$\left.\begin{aligned} \xi_j &= x_j \quad (j = 0, \dots, p), \\ \epsilon &= y - \beta_0 - \beta_1 x_1 - \dots - \beta_p x_p. \end{aligned}\right\} \tag{30}$$

The transformation has a unit Jacobian and if $f(y, x_1 \dots, x_p)$ is the frequency function of y and the x's we find, for the probability of the sample, a product of n terms like

$$f(\epsilon + \sum_{j=0}^{p} \beta_j \xi_j, \xi_1, \dots, \xi_n)\, d\epsilon\, d\xi_1 \dots d\xi_n. \tag{31}$$

To treat the x's as determined variates we shall require a factorization

$$f(\epsilon+\Sigma\beta_j\xi_j,\xi_1,\ldots,\xi_n)=g(\xi_1,\ldots,\xi_n)\,h(\epsilon), \tag{32}$$

which is equivalent to $$f(y,x_1,\ldots,x_p)=g(x_1,\ldots,x_p)\,h(y-\Sigma\beta_j x_j). \tag{33}$$

We find a similar factorization of the characteristic functions

$$\phi(u,t_1,\ldots,t_p)=e^{iu\beta_0}\phi_g(t_1+\beta_1u,\ldots,t_p+\beta_pu)\,\phi_h(u), \tag{34}$$

where u refers to y and t_j to x_j.

In terms of cumulants we have

$$\Sigma\frac{\kappa_{r,s_1,\ldots,s_p}(iu)^r(it_1)^{s_1}\ldots(it_p)^{s_p}}{r!\,s_1!\ldots s_p!}=iu\beta_0+\Sigma\frac{\kappa'_{s_1,\ldots,s_p}i^{s_1+\cdots+s_p}(t_1+\beta_1u)^{s_1}\ldots(t_p+\beta_pu)^{s_p}}{s_1!\ldots s_p!}+\Sigma\frac{\kappa''_r(iu)^r}{r!}. \tag{35}$$

Considering those terms where u appears only as the first power and taking $\kappa''_1=0$ we find

$$\kappa_{1,s_1,\ldots,s_p}=\beta_1\kappa'_{s_1+1,s_2,\ldots,s_p}+\beta_2\kappa'_{s_1,s_2+1,\ldots,s_p}+\cdots+\beta_p\kappa'_{s_1,\ldots,s_p+1} \tag{36}$$

as a generalization of (19). In particular, taking $s_j=1$ and all the other s's as zero, we have

$$\kappa_{1,0,\ldots1,\ldots,0}=\beta_1\kappa'_{1,0,\ldots,1,\ldots,0}+\beta_2\kappa_{0,\ldots,2,\ldots,0}+\cdots+\beta_p\kappa_{0,\ldots,1,\ldots,1}, \tag{37}$$

which are the standard equations (29) for the parent population and show that in any case equations of type (29) are consistent.

24. To give the estimation equations (29) more than consistency we shall again require assumptions of normality of the usual type. From (23) we see that h continues to embody all the available information about the β's, and a maximum likelihood solution of type (29) can be produced from it if we require that ϵ shall be a normal variate with variance σ^2 independently of the x's and that the various values of ϵ are independent. As in the case of one predicated variate, we may take the predicated variates as determined if and only if the regression is linear; and to produce the estimating equations we require the residual to be normal. The application of the t-test follows under the usual assumptions.

25. We next remove the restriction as to linearity by observing that $x_1,\ldots,x_p$ may in particular cases be functionally dependent and could, for example, be the powers of a single variate $x_1, x_1^2\ldots x_1^p$. Clearly, however, in such a case, by determining one variate we determine them all. This extension does not contradict the statement in §15 that we may regard the undetermined case as equivalent to the determined case only if the regression is linear. What §15 implies is that if we are estimating the constant β_1 in the equation $Y=\beta_0+\beta_1X$ we may regard x as a determined variate only if the parent regression really is linear. If it really is curvilinear, of the form, say,

$$Y=\beta_0+\beta_1X+\beta_2X^2,$$

then we may regard x as a determined variate only if we estimate β_1 and β_2 together.

26. Up to this point nothing particularly new has emerged from the treatment except perhaps some clarification and the extensions of equation (36). I shall resume the line of development later in Part II of the paper in a discussion of structure in stochastic systems. At this stage I turn to a consideration of the problem of functional relationship and the method of attack developed by Frisch and Reiersol; but before doing so I raise one point of theoretical importance in connexion with significance tests of regression coefficients.

27. When we are fitting regressions (say, for simplicity, with one predicated variate) we usually begin with simple linearity and proceed to take in quadratic, cubic and more complicated terms as necessity arises. The criterion of 'necessity' in this connexion is based on the size of the residual variance. One popular method of applying it is to test the significance of successive coefficients b_2, b_3, etc., as they are added and to stop at the point when one of them fails to attain significance. Now in view of § 25 above I doubt whether this is legitimate in the undetermined case. Suppose a regression line is really quadratic. We calculate a simple linear regression giving an observed coefficient b_1. Now we can only regard the undetermined case as equivalent to the determined case if the parent regression really is linear, which it is not. The least-squares method of estimation and the t-test of significance must then fail or, at least, lose their previous theoretical justification. The estimator will be biased and the power of the test unknown.

Or to put it another way: we set up the provisional hypothesis that the regression is linear. We find a value of b_1 which is not significant on an ordinary t-test. What should be concluded? The usual conclusion would be that y and x can be regarded as independent, or rather that the mean of y is constant for all x. But it would be equally valid to conclude that relationship may exist in a non-linear form. We arrive at the basic contention of the Neyman-Pearson theory, that a test should specify the alternative hypothesis.

28. So far I have followed the classical line of development to the extent of assuming that, both in the determined and in the undetermined case, the variates are measurable without error. This is rather exceptionally so in practice, although cases often arise where errors of observation are negligible compared with the actual variation among individuals. If we wish to take account of random errors of observation an essentially new situation arises. In fact, several new situations arise, and it is necessary to distinguish between them with some care.

29. Consider in the first place the undetermined case of variates x, y distributed with frequency function $f(x, y)$. We suppose that x is subject to an experimental error v and y to an experimental error w. Thus our observed variables x' and y' are

$$\left.\begin{aligned} x' &= x+v, \\ y' &= y+w. \end{aligned}\right\} \tag{38}$$

We meet here with a new kind of logical difficulty. Unless our technique of measurement can be pushed to the point of perfection we can never observe either x or v. Only the values x' appear. Any hypotheses we make about x and v are thus capable of only very indirect verification, because they are so difficult to disentangle. In such circumstances it is natural to inquire why we do not work directly with the observables x' and y' and regard them as our primary variables, treating v and w as part of the obscure complex of causation which gives rise to the observed variation—'washing them into the error term'. Sometimes we may in fact do so; but more frequently we desire to separate the error of observation from the other causes of variation, e.g. because we want to make separate hypotheses about its behaviour, or because it is a special cause of variation which we hope to be able to reduce.

30. To make much progress we must make some hypothesis about the behaviour of v and w. It is usually reasonable to suppose that each of v and w is independent of each of x and y and, perhaps, that v and w are independent of each other. In regression analysis we may then be interested in the regression of y' on x, of y' on x', of y on x and of y on x'. In fact,

there are four regression lines which coincide in pairs. This part of the subject has been treated fully in an admirable review by Lindley (1947), who shows, among other things, that the regression of y' on x' is not linear even if that of y on x is linear unless certain conditions are obeyed by the cumulants of the distribution of x and y. In particular, if the distribution of x is normal only normality in v will preserve the linearity of y' on x'.

31. Let us now consider whether we can have a situation corresponding to the determined case when the predicated variate is subject to error. Obviously we cannot do so in any strict sense, for we cannot fix the value of v in any given case. We can, so to speak, aim at a particular value of x but not with any assurance of reaching that precise value. Thus our solution of the values of the predicated variate is essentially undetermined. To what extent, if any, can we regard it as determined for the purposes of estimation or significance tests?

Something depends here on what we are trying to estimate. We may be interested in the regression of y on x as involving some inherent relationship irrespective of the observational errors; or we may require the regression of y on x' because we intend to use the equation for prediction and can observe only x'. I consider these two cases in turn.

32. If the observed variates x', y' are distributed as $f(x', y')$ and this may be written as $f_1(x, y) f_2(v) f_3(w)$, then the foregoing argument applies unchanged to show that the maximum likelihood estimator of the constant β_1 in $Y = \beta_0 + \beta_1 X$ is given by

$$\Sigma yx - b_1 \Sigma x^2 = 0. \tag{39}$$

This obviously must be so, for we are merely ignoring the errors v and w by factorizing them from the frequency function. But the quantities occurring in (39) are not what we observe. If we substitute from (38) we find

$$\Sigma(y' - w)(x - v) - b_1 \Sigma(x' - v)^2 = 0,$$

or

$$\Sigma y'x' - \Sigma wx' - \Sigma vy' + \Sigma vw - b_1(\Sigma x'^2 - 2\Sigma x'v + \Sigma v^2) = 0, \tag{40}$$

Only two of the seven quantities required to solve this equation are observed. To go further we must argue from the particular to the general. We know from our hypothesis about v and w that the mean values of terms like $\Sigma wx'$ and Σwv vanish. As an unbiased estimating equation we then have

$$\Sigma y'x' - b_1\{\Sigma x'^2 + \Sigma v^2\} = 0. \tag{41}$$

This is much better, but there remains the term Σv^2, proportional on the average to $(n-1)\sigma^2/n$, where σ^2 is the variance of v. We meet here, as elsewhere throughout the theory of component analysis, the simple but important fact that we cannot estimate the relationship $Y = \beta_0 + \beta_1 X$ *with any known accuracy whatsoever* unless we know the variance of the errors of observation.* This is obvious enough, but it is a point worth bringing into prominence, because in the more usual type of statistical inquiry we can estimate the error of observation from the data themselves, whereas here we cannot. We must assume something about it or try to determine it from collateral experience.

33. It should also be noticed that although (41) is an unbiased estimating equation the derived estimator of β_1, namely,

$$b_1 = \frac{\Sigma y'x'}{\Sigma x'^2 + \Sigma v^2}, \tag{42}$$

is not necessarily unbiased even if we know Σv^2. To some extent this is due to our rather arbitrary definition of bias. We must draw a distinction between an unbiased estimator and an unbiased estimating equation.

* If Y is a lagged value of X or vice versa the above statement ceases to be true.

34. The estimator b_1 given by (42) is clearly not distributed as Student's t, and in the general case I do not know how to test its significance. Even if we assume the distribution of v known (with known variance) the problem of significance is formidable. It is not, however, inaccessible to inquiry—for example, the approximative method used by Welch (1938) on the Behrens test might be applicable. One thing seems clear: we cannot expect the t-distribution, or any other distribution, to hold unless we fix both x' and v, which only brings us back to the case when effectively we can ignore the error of observation. It appears, therefore, that the determined case cannot arise.

35. Now consider the case when we are interested in determining a regression equation of type
$$Y = \beta_0 + \beta_1' X', \tag{43}$$
i.e. the regression of y on the observed x'. The same type of argument which we have already employed leads to the estimating equation
$$\Sigma yx' - b_1' \Sigma x'^2 = 0, \tag{44}$$
or
$$\Sigma y'x' - \Sigma vx' - b_1' \Sigma x'^2 = 0. \tag{45}$$
Assuming that v is independent of x we have the unbiased estimating equation
$$\Sigma y'x' - b_1' \Sigma x'^2 = 0, \tag{46}$$
which depends only on the observables x' and y'. We do not require the variances of v and w.

The t-test of the estimator b may be applied only if we assume that y' and x' are distributed in such a way that
$$f(x', y') = g(x')\, h(x' - \beta_0 - \beta_1' y'), \tag{47}$$
and that h is normal.

36. In §§ 32–35 I have passed over one point in order to arrive at the estimating equations, and I now return to it. In the case leading to (39) we are assuming that for the variates x and y there is a basic distribution such that we may write
$$y_X = \beta_0 + \beta_1 X + \epsilon, \tag{48}$$
and take ϵ to obey the usual assumptions of normality. Otherwise the estimating equation (39) is not warranted by maximum likelihood. To justify (44) in the same way we require to be able to write
$$y_{X'} = \beta_0' + \beta_1' X' + \epsilon', \tag{49}$$
where again ϵ' obeys the normality conditions. Now (48) and (49) are not in general compatible. Suppose in fact that the distribution of x, y is given by $g(x)\, h(y - \beta_0 - \beta_1 x)$ with characteristic function $\phi_g(t_1 + t_2\beta_1)\, \phi_h(t)\, e^{it_2\beta_0}$. The characteristic function of x' and y' is then
$$e^{it_2\beta_0}\, \phi_g(t_1 + t_2\beta_1)\, \phi_h(t_2)\, \phi_v(t_1)\, \phi_w(t_2),$$
and this must be equal to, say, $e^{it_2\beta_0'}\, \phi_G(t_1 + t_2\beta_1')\, \phi_H(t_2)$.

Thus
$$\frac{\phi_g(t_1 + t_2\beta_1)\, \phi_v(t_1)}{\phi_G(t_1 + t_2\beta_1')} = \frac{\phi_H(t_2)}{\phi_h(t_2)\, \phi_w(t_2)}\, e^{-it_2(\beta_0' - \beta_0)}$$
$$= \text{function independent of } t_1$$
$$= \frac{\phi_g(t_2\beta_1)}{\phi_G(t_2\beta_1')}$$
as follows from putting $t_1 = 0$ on the left-hand side. Hence
$$\frac{\phi_g(t_1 + t_2\beta_1)}{\phi_g(t_2\beta_1)}\, \phi_v(t_1) = \frac{\phi_G(t_1 + t_2\beta_1')}{\phi_G(t_2\beta_1')}. \tag{50}$$

In terms of cumulants, writing κ_r^g for the rth cumulant of g, etc. and, absorbing i into t temporarily, we then have

$$\Sigma \frac{\kappa_r^g(t_1+t_2\beta_1)^r}{r!} - \Sigma \frac{\kappa_r^g(t_2\beta_1)^r}{r!} + \Sigma \kappa_r^v \frac{t_1^r}{r!} = \Sigma \frac{\kappa_r^G(t_1+t_2\beta_1')^r}{r!} - \Sigma \frac{\kappa_r^G(t_2\beta_1')^r}{r!}. \tag{51}$$

The cumulants κ_1 merely locate means, the terms of second power give us

$$\kappa_2^g + \kappa_2^v = \kappa_2^G, \tag{52}$$

$$\beta_1 \kappa_2^g = \beta_1' \kappa_2^G. \tag{53}$$

These are equivalent to Lindley's equation (7). Consider now a term of order $r > 2$. The identification in powers of t_1 and t_2 give us r equations, $r-1$ of which concern only κ_r^g and κ_r^G. If β_1 and β_1' are not zero and are not equal this entails the vanishing of those cumulants.

Thus, for variates with existing variances:

If (48) and (49) are to hold for two different non-vanishing values β_1 and β_1' the distributions $g(x)$, $G(x)$, and $f(v)$ are all normal.

And if the residual ϵ is normal, y and x are jointly normally distributed.

37. Theorems in which linearity of regression implied normality have been proved by Allen (1938) and Lindley (1947). Allen proved that if

$$\left.\begin{aligned} x' &= lx + v, \\ y' &= mx + w, \end{aligned}\right\} \tag{54}$$

then the necessary and sufficient condition for the regression of y' and x' to be linear for all l and m is that x and v are normally distributed. She required in her proof that all moments of x and v should exist, but Lindley showed that it is enough to assume the variance of x finite.*

Lindley proved a general theorem for any number of predicated variates which in the case of a single variate reduces to this:

If

$$x' = x + v,$$

$$y' = y + w,$$

and v, w are independent random variates which are independent of x and y, then the necessary and sufficient condition that the regression of y' on x', say $Y = \beta_0' + \beta_1' X'$, is linear, if that of y on x, say $Y = \beta_0 + \beta_1 X$, is linear, is

$$\beta_1' \psi_{x'} = \beta_1 \psi_x, \tag{55}$$

where ψ is the cumulative function (the logarithm of the characteristic function). This may be put in the equivalent form

$$(\beta_1' - \beta_1) \frac{d\psi_{x'}}{dt} = \beta_1 \frac{d\psi_v}{dt}. \tag{56}$$

The reason that the theorem of §36 goes further than Lindley's result (though consistent with it) is that I am assuming more than linearity of regression. He requires only linearity; I require the residual ϵ' to be independent of X'. The effect of my theorem is that I can only get it if the distributions of x and v are normal.

38. The effect of such results on the investigation of regression lines in practice is clearly important. The errors of observation impair our estimators, vitiate our tests of significance and even bend our regression lines unless we are prepared to postulate normality in the

* Rao (1947, and correction, 1949) and Fix (1949) have relaxed the conditions still further.

variates, which is asking rather a lot, particularly in economic work. But they are not quite as devastating as they look. A slight departure from linearity will sometimes allow the ordinary theory to be used as an approximation. None the less, it would appear that much more effort is needed to reduce errors of observation than has sometimes been supposed. We are, I hope, moving out of the phase in statistical development when we gloried in a profusion of errors because of a pious hope in the central limit effect and the supposition that everything unknown could be swept into a single error term.

39. I turn to the subject of functional relationships between variables. Let us begin by considering the problem of estimating the 'true' linear relation between two variables X and Y, both subject to errors of observation. This is the situation common in physics where we have, say, a number of observations on the length and temperature of a metal bar and wish to estimate the constant α in the relation $L = \alpha T$, or in economics where we postulate a relation between demand D and price P

$$\log P = \alpha \log D.$$

In such a case we are *not* concerned with a parental population of values of random variates x and y, l and t, or d and p. As a rule one of our variables at least can be arranged to have approximately any value we like (though this is rarely so in economics) with any frequency we like. The problem is no longer to estimate a regression coefficient. It is to estimate the constants in a functional relation

$$Y = \alpha_0 + \alpha_1 X. \tag{57}$$

40. The fundamental difficulty here is that we cannot replicate our observations by fixing values of X or Y. We observe two quantities x' and y' given by

$$\left.\begin{aligned} x'_X &= X + v_X, \\ y'_Y &= Y + w_Y, \end{aligned}\right\} \tag{58}$$

but as v and w are random variables we can never estimate the corresponding value of X and Y even if we make the usual assumptions that v_X does not depend on X, w_Y on Y and that they are independent of each other. We have from (58) and (57)

$$(y'_Y - w_Y) = \alpha_0 + \alpha_1 (x'_X - v_X), \tag{59}$$

and our random variables y' and x' are dependent on parameters Y and X and hence may vary from one observation to another, so that we never know whether we are observing the same distribution; and in fact we can safely assume in most cases that we are not. Equation (59) is equivalent to

$$y'_Y = \alpha_0 + w_Y - \alpha_1 v_X - \alpha_1 x'_X, \tag{60}$$

and if we want to estimate α_0 and α_1 from the observed y' and x' we have to make some assumption about the 'residual' $\alpha_0 - w_Y - \alpha_1 v_X$, which depends on α_0 and α_1. Or to put it another way, we require to be able to estimate not only α_0 and α_1 but X and Y for each observation. This brings us back to the case of essential indeterminacy in component analysis which I mentioned in § 21. In fact, our present problem is one of the simpler problems in component analysis.

41. What, then, are we to do? It is clearly no use telling the physicist that his problem is insoluble. He solves it often enough on intuitive grounds. What we can fairly tell him is that it is insoluble without some further information about the two error distributions of v and w; and in fact all that we need to know is the ratio of variance of v and w.

For the observed sample of n values of x and y we have the likelihood function

$$\frac{1}{(2\pi)^{\frac{1}{2}n}(\operatorname{var} v \operatorname{var} w)^{\frac{1}{2}n}} \exp\left\{-\frac{\Sigma(x'_X - X)^2}{2 \operatorname{var} v}\right\} \exp\left\{-\frac{\Sigma(y'_Y - Y)^2}{2 \operatorname{var} w}\right\}$$

or, say,

$$\frac{1}{(2\pi)^{\frac{1}{2}n}(\sigma_v \sigma_w)^n} \exp -\frac{1}{2}\left[\frac{\Sigma(x'_X - X)^2}{\sigma_v^2} + \frac{\Sigma(y'_Y - \alpha_0 - \alpha_1 X)^2}{\sigma_w^2}\right].$$

But if we maximize this likelihood for variations in $X_1, \ldots, X_n, \sigma_v$ and σ_w, we get an unacceptable result (cf. Lindley, 1947, p. 235 and Kendall, 1950, p. 80), namely, that $\sigma_v = \sigma_w$. We must, therefore, introduce a further assumption. Lindley remarks that the most common assumption is that the ratio σ_w/σ_v is known. This produces consistent estimates of α_0 and α_1.

42. I leave the question of functional relationship at this point. It cannot be said that the problems have been satisfactorily solved. The method of maximum likelihood seems to break down, and we require a new principle of inference or a new assumption concerning the error variance before we can make progress. It may well be that different problems will require different assumptions so that no general solution exists, or it may be that we can only estimate limits to the error variances and hence that a solution of the problem is possible only in approximate terms.

43. I had intended to include in this paper an examination of some questions arising on structural equations affected by stochastic elements. There has recently appeared a very important monograph from the Cowles Commission (Koopmans, 1950) which brings together a great deal of the work done by econometric writers on this subject. It settles some difficulties, but raises many others, and my treatment will have to be more extended than was originally proposed. I therefore intend to return to the matter in a separate Part II of this paper which will appear at a later date.

REFERENCES

ALLEN, H. V. (1938). A theorem concerning the linearity of regression. *Stat. Res. Mem.* **2**, 60.

FISHER, R. A. (1922). The goodness of fit of regression formulae and the distribution of regression coefficients. *J. R. Statist. Soc.* **85**, 97.

FIX, EVELYN (1949). Distributions which lead to linear regressions. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press.

KENDALL, M. G. (1948). *The Advanced Theory of Statistics*, **1**, 4th ed. London: Charles Griffin and Co.

KENDALL, M. G. (1950). Factor analysis as a statistical technique. *J.R. Statist. Soc.* B, **12**, 60.

KOOPMANS, T. C. (editor) (1950). *Statistical Inference in Dynamic Economic Models*. New York: John Wiley and Sons; London: Chapman and Hall.

LINDLEY, D. V. (1947). Regression lines and the linear functional relationship. *J.R. Statist. Soc. Suppl.* **9**, 218.

RAO, C. R. (1947, 1949). Note on a problem of Ragnar Frisch. *Econometrica*, **15**, 245 and Correction, *Econometrica*, **17**, 212.

WELCH, B. Z. (1938). The significance of the differences of two means when the population variances are unequal. *Biometrika*, **29**, 350.

PARTIAL AND MULTIPLE RANK CORRELATION

By P. A. P. MORAN, *Institute of Statistics, University of Oxford*

Apart from a paper by M. G. Kendall (1942), partial rank correlation has been very little considered and the analogous definitions of multiple rank correlation have not been considered at all. The present paper considers some of the problems which arise in the study of such coefficients.

By using the generalized method of defining coefficients of rank correlation, given by Daniels (1944), Spearman's coefficient of rank correlation r_s, and Kendall's t can both be shown to be product-moment correlation coefficients between sets of suitably defined scores a_{ij} attached to each pair of elements in the ranking. Whatever the probability distribution assumed to account for the ranking, these scores are not independently distributed and this fact bedevils the resulting distribution theory of the coefficients. If all the $n!$ permutations of the ranks are equiprobable, r_s has a variance $1/(n-1)$ and t a variance $2(2n+5)/9n(n-1)$. They are both distributed symmetrically between the limits ± 1. Kendall has shown that the distribution of r_s is well approximated by a Beta-type distribution so that

$$r_s\left(\frac{n-2}{1-r_s^2}\right)^{\frac{1}{2}} \tag{1}$$

has Student's t-distribution, with $n-2$ degrees of freedom, to a high degree of approximation.

It is perhaps worth pointing out that the same method gives a good approximation to a test for t and that, analogously to r_s, the distribution of

$$t\left\{\frac{\frac{9n(n-1)}{2(2n+5)}-1}{1-t^2}\right\}^{\frac{1}{2}} \tag{2}$$

where t is Kendall's t, is well approximated by Student's t-distribution with $\dfrac{9n(n-1)}{2(2n+5)}-1$ 'degrees of freedom'. The use of this in practice may involve interpolation for fractional 'degrees of freedom'. For $n \geqslant 10$ Kendall has shown that the normal approximation is satisfactory.

In order to test how good an approximation is given by the above procedure, the exact probabilities for the tails of the distribution of $S = \frac{1}{2}n(n-1)t$ given by Kendall (1948) for $n = 8, 9, 10$ were compared with values given by testing t as if it were an ordinary product-moment correlation coefficient based on $1+\{9n(n-1)/2(2n+5)\}$ pairs of observations. Subtracting unity from S to obtain a continuity correction, dividing by $\frac{1}{2}n(n-1)$, and doing a double interpolation in Miss David's tables (1938) of the Correlation Coefficient we obtain the probability levels given in Table 1 below under the headings 'Approx.' There is a possible rounding error of unity in the last decimal place of these probabilities. It is clear that agreement is good and the test can be safely used for $n \geqslant 8$. A somewhat better fit would be obtained by dividing S by $\frac{1}{2}n(n-1)+2$ instead of $\frac{1}{2}n(n-1)$. The formulae are then rather more complicated.

The fact that t has a considerably smaller variance than r_s does not imply that it provides a more powerful test of independence than r_s any more than the fact that

$$\text{var}\,(r_s) = 1/(n-1) = \text{var (product-moment correlation coefficient, } r)$$

Table 1. *Probabilities that $S \geqslant k$ for $n = 8, 9, 10$*

	$n=8$		$n=9$			$n=10$	
k	Exact	Approx.	Exact	Approx.	k	Exact	Approx.
14	0·054	0·055	—	—	19	0·054	0·055
16	0·031	0·030	0·060	0·060	21	0·036	0·036
18	0·016	0·014	0·038	0·037	23	0·023	0·023
20	—	—	0·022	0·021	25	0·014	0·013
22	—	—	0·012	0·011			

implies that a test based on r_s is as powerful as one based on r. If we take as our alternative hypothesis that the ranks are derived by ranking a sample of n pairs of observations (x_i, y_i) from a bivariate normal distribution with $\rho \neq 0$, the relative power of the three tests for large n can be gauged by their efficiency in estimating ρ, since we know their expectations and variance. For in this case we have (see e.g. Kendall (1949))

$$E(t) = \frac{2}{\pi}\sin^{-1}\rho, \qquad \text{(Greiner)} \quad (3)$$

$$\text{var}\,(t) = \frac{2}{n(n-1)}\left\{1 - \left(\frac{2}{\pi}\sin^{-1}\rho\right)^2 + 2(n-2)\left[\frac{1}{9} - \left(\frac{2}{\pi}\sin^{-1}\tfrac{1}{2}\rho\right)^2\right]\right\}, \qquad \text{(Esscher)} \quad (4)$$

$$E(r_s) = \frac{6}{\pi(n+1)}\{\sin^{-1}\rho + (n-2)\sin^{-1}\tfrac{1}{2}\rho\}, \qquad \text{(Moran)} \quad (5)$$

$$\text{var}\,(r_s) = \frac{1}{n}\{1 - 1{\cdot}563465\rho^2 + \ldots\} + O(n^{-2}). \qquad \text{(Kendall)} \quad (6)$$

If n is large and ρ small we therefore have

$$E(t) \doteqdot \frac{2}{\pi}\rho, \quad \text{var}\,(t) \doteqdot \frac{4}{9n},$$

$$E(r_s) \doteqdot \frac{3}{\pi}\rho, \quad \text{var}\,(r_s) \doteqdot \frac{1}{n},$$

and therefore

$$\frac{E(t)}{\text{s.e.}\,(t)} \doteqdot \frac{3n^{\frac{1}{2}}\rho}{\pi} \doteqdot \frac{E(r_s)}{\text{s.e.}\,(r_s)}.$$

It therefore follows that, under these assumptions, t and r_s are asymptotically equally efficient in estimating ρ when ρ is small and equally powerful in testing the hypothesis that $\rho = 0$. This result is not surprising when we recall that in the universe of equiprobable permutations the correlation of t and r_s tends to unity very rapidly as n increases (Kendall *et al.* (1939)), so that for most rankings, t is about $\frac{2}{3}$ of r_s. It has been recently shown (David, Kendall & Stuart, 1951) that t and r_s are also highly correlated when calculated from rankings derived from bivariate normal distributions.

Hotelling and Pabst (1936) have shown that when used to provide an estimator of ρ, r_s has an efficiency asymptotically equal to $9\pi^{-2} = 0{\cdot}91$, so that it follows that the use of t and r_s involves a loss of about 9 % of the information. K. Pearson also regarded $9\pi^{-2}$ as a measure of efficiency, but approached the problem from the point of view of 'grades'.

Partial t

Given three rankings, which we denote by 1, 2 and 3, Kendall (1942) defines a partial ranking correlation coefficient $t_{12.3}$ in such a way that

$$t_{12.3} = \frac{t_{12} - t_{13}t_{23}}{(1-t_{13}^2)^{\frac{1}{2}}(1-t_{23}^2)^{\frac{1}{2}}}. \tag{7}$$

In fact, $t_{12.3}$ is the partial correlation between the sets of t-scores. Very little is known about the distribution of $t_{12.3}$. Hoeffding (1948) has shown that if $\tau_{12}, \tau_{13}, \tau_{23}$ are defined for a continuous population and the corresponding t's obtained by ranking a sample from this population, then the distribution of $n^{\frac{1}{2}}(t_{12.3} - \tau_{12.3})$ tends to normality with zero mean, and a variance which he evaluates, provided $\tau_{13}^2 \neq 1$, $\tau_{23}^2 \neq 1$. If in addition $\tau_{13} = \tau_{23} = 0$, the distribution of $n^{\frac{1}{2}}(t_{12.3} - \tau_{12.3})$ is asymptotically the same as that of $n^{\frac{1}{2}}(t_{12} - \tau_{12})$.

It seems very difficult to prove anything about the distribution of $t_{12.3}$ for small n. We first notice that we must eliminate from our consideration all those sets of rankings for which $t_{13} = 1$ or $t_{23} = 1$ for in such cases $t_{12.3}$ is undefined. In order to see what happens in practice I calculated (7) in the case where $n = 4$, ranking 3 is (1234) and rankings 1 and 2 have all permutations of (1234) except (1234) and (4321). There are $(4!-2)^2$ of these, but the results are symmetrical in rankings 1 and 2 and the coefficient reverses its sign when either of the latter are reversed. It follows that only about $\frac{1}{8}(4!-2)^2$ calculations are necessary. The result is shown in Table 2, which gives values of $t_{12.3}$ for positive values of t_{13} and t_{23}.

Table 2. *Values of* $t_{12.3}$

	Ranking 1 t_{13}	2134 0·6667	1324 0·6667	1243 0·6667	2314 0·3333	2143 0·3333	3124 0·3333	1342 0·3333	1423 0·3333	2314 0	2341 0	241 0
Ranking 2	t_{23}											
2134	0·6667	1·0000	−0·2000	−0·2000	0·6325	0·6325	−0·3162	−0·3162	−0·3162	0·4472	0·4472	0·44
1324	0·6667	−0·2000	1·0000	−0·2000	−0·3162	−0·3162	−0·6325	0·6325	−0·3162	0·4472	−0·4472	−0·44
1243	0·6667	−0·2000	−0·2000	1·0000	−0·3162	0·6325	−0·3162	−0·3162	0·6325	−0·4472	−0·4472	0·44
2314	0·3333	0·6325	−0·3162	−0·3162	1·0000	0·2500	0·2500	−0·5000	−0·5000	0·7071	0·7071	0
2143	0·3333	0·6325	−0·3162	0·6325	0·2500	1·0000	−0·5000	−0·5000	−0·5000	0	0	0·70
3124	0·3333	−0·3162	0·6325	−0·3162	0·2500	−0·5000	1·0000	0·2500	−0·5000	0·7071	0	−0·70
1342	0·3333	−0·3162	0·6325	−0·3162	−0·5000	−0·5000	0·2500	1·0000	0·2500	0	−0·7071	−0·70
1423	0·3333	−0·3162	−0·3162	0·6325	−0·5000	−0·5000	−0·5000	0·2500	1·0000	−0·7071	−0·7071	0
3214	0	0·4472	0·4472	−0·4472	0·7071	0	0·7071	0	−0·7071	1·0000	0·3333	−0·33
2341	0	0·4472	−0·4472	−0·4472	0·7071	0	0	−0·7071	−0·7071	0·3333	1·0000	0·33
2413	0	0·4472	−0·4472	0·4472	0	0·7071	−0·7071	−0·7071	0	−0·3333	0·3333	1·00

From this it is clear that, unlike the case of product moment partial correlation, $E(t_{12.3})$, when t_{13} and t_{23} are kept fixed, is not necessarily equal to zero. This is also true for other values of n. Suppose rankings 1 and 2 are each obtained from 3 by a single interchange of nearest neighbours, then t_{13} and t_{23} are each equal to $1 - \dfrac{2}{n(n-1)}$, and t_{12} is equal to 1 (with probability $(n-1)^{-1}$) or equal to $1 - \dfrac{4}{n(n-1)}$ (with probability $(n-2)/(n-1)$). It will then be found that the expectation of the numerator of $t_{12.3}$ is not equal to zero.

I have not been able to find a formula either for $E(t_{12.3} \mid t_{13}, t_{23})$ or for $\operatorname{var}(t_{12.3})$ when t_{13}, t_{23} are fixed or variable. Since t_{12} is approximately distributed as r based on $\dfrac{9n(n-1)}{2(2n+5)} + 1$ pairs of observations, it is natural to suppose that $t_{12.3}$ is approximately distributed as

r based on $\frac{9n(n-1)}{2(2n+5)}$ pairs of observations. If we keep t_{13} and t_{23} fixed, only a few values of $t_{12.3}$ are possible when $n = 4$ and we therefore take the case where rankings 1 and 2 range over all possible permutations except (1234) and (4321). The frequency distribution of $t_{12.3}$ is then given in Table 3 which is obtained from Table 2 when completed by reflexion. As the distribution of $t_{12.3}$ is symmetric, the frequencies are only given for the positive values of $t_{12.3}$ together with the whole frequency for $t_{12.3} = 0$.

Table 3

$t_{12.3}$	0	0·2000	0·2500	0·3162	0·3333	0·4472	0·5000	0·6325	0·7071	1·0000
Frequency	24	6	8	18	6	18	12	12	18	11

The variance of this distribution is $0{\cdot}282 = (3{\cdot}55)^{-1}$. On the other hand $9n(n-1)/2(2n+5)$ for $n = 4$ is equal to 4·154 so that agreement is not very good. One source of discrepancy arises from the fact that $2(2n+5)/9n(n-1)$ is the variance of t over all permutations and in the above we have omitted the rankings (1234) and (4321). If we make a correction for this, however, the discrepancy is increased. It is unfortunate that calculations of the above kind quickly become impracticable when n increases. $n = 5$, for example, would require approximately $5^2 = 25$ times as much calculation as $n = 4$. It therefore seems that no very clear conclusions about the distribution of partial t can be deduced from the above discussion.

MULTIPLE RANK CORRELATION

It is, however, possible to make somewhat more progress with the theory of multiple rank correlation. Taking first a definition based on t-scores, we write ${}^tR^2_{1(23)}$ for the square of the multiple t-rank correlation coefficient between a ranking 1 and two rankings 2, 3 and we define it by the equation

$$1 - {}^tR^2_{1(23)} = (1 - t^2_{13})(1 - t^2_{12.3}). \tag{8}$$

${}^tR^2_{1(23)}$ is, in fact, the square of the product moment multiple correlation coefficient between the set of scores for the ranking 1 and the sets for rankings 2 and 3.

Inserting the value of $t_{12.3}$ in terms of t_{12}, t_{13} and t_{23}, we find

$${}^tR^2_{1(23)} = \frac{t^2_{12} + t^2_{13} - 2t_{12}t_{23}t_{13}}{(1 - t^2_{23})}. \tag{9}$$

We now find the expectation of (9) when rankings 2 and 3 are kept fixed and ranking 1 ranges, with equal probability, over all possible permutations of $(1, \ldots, n)$. In terms of this expectation we can then determine an approximate distribution for ${}^tR^2_{1(23)}$. From (9) it follows that

$$E({}^tR^2_{1(23)}) = \frac{2}{(1 - t^2_{23})}\{E(t^2_{13}) - t_{23}E(t_{12}t_{13} \mid t_{23})\}$$

$$= \frac{2}{(1 - t^2_{23})}\left\{\frac{2(2n+5)}{9n(n-1)} - t_{23}E(t_{12}t_{13} \mid t_{23})\right\}. \tag{10}$$

Take ranking 3 to be in the natural order $1, 2, \ldots, n$. Let a_{ij}, b_{ij} $(i < j)$ be the scores for the rankings 1 and 2 relative to 3. Then $a_{ij}b_{ij}$ will be the score of 2 relative to 1. To evaluate the second term inside the bracket in (10) we have to find

$$t_{23}\left(\frac{2}{n(n-1)}\right)^2 E\{\sum_{i<j} a_{ij}b_{ij} \sum_{p<q} a_{pq}\},$$

when ranking 1 has equal probability for all its $n!$ possible permutations and rankings 2 and 3 are kept fixed. Then

$$E\{(\Sigma a_{ij}b_{ij})(\Sigma a_{pq})\} = \Sigma a_{ij}^2 b_{ij} + \Sigma a_{ij}a_{ik}b_{ij} + \Sigma a_{ij}a_{jk}b_{ij} + \Sigma a_{ij}a_{ki}b_{ij} + \Sigma a_{ij}a_{hj}b_{ij} + \Sigma a_{ij}a_{lm}b_{ij}$$
$$= c_1 + c_2 + c_3 + c_4 + c_5 + c_6, \quad \text{say.}$$

In the right-hand side of the above expression, different letters used for suffixes are to be understood as denoting different integers. We now find each of these terms separately. Clearly

$$c_1 = E(\Sigma a_{ij}^2 b_{ij}) = \Sigma b_{ij}.$$

To find the next four terms, consider the six possible permutations of three objects. By making a table of the corresponding scores we easily see that

$$E(a_{ij}a_{ik}) = E(a_{ij}a_{hj}) = \tfrac{1}{3}.$$
$$E(a_{ij}a_{jk}) = E(a_{ij}a_{hi}) = -\tfrac{1}{3}.$$

Given i and j, there are $n-i-1$ possible values for k in the sum $\Sigma a_{ij}a_{ik}b_{ij}$ and therefore $c_2 = \frac{1}{3}\sum_{i<j}(n-i-1)b_{ij}$. Similarly we find

$$c_3 = -\tfrac{1}{3}\sum_{i<j}(n-j)b_{ij}.$$
$$c_4 = -\tfrac{1}{3}\sum_{i<j}(i-1)b_{ij}.$$
$$c_5 = \tfrac{1}{3}\sum_{i<j}(j-2)b_{ij}.$$

Finally it is obvious that $c_6 = 0$. Adding we get

$$E\{(\Sigma a_{ij}b_{ij})(\Sigma a_{pq})\} = \tfrac{1}{3}\sum_{i<j} b_{ij}(1+2j-2i).$$

In the above system of scoring we have only considered scores a_{ij}, b_{ij} for $i<j$. We now revert to Daniels's original scoring system and write

$$\beta_{ij} = b_{ij} \quad \text{if} \quad i<j,$$
$$= -b_{ji} \quad \text{if} \quad i>j,$$
$$= 0 \quad \text{if} \quad i=j.$$

Then

$$2\sum_{i<j} b_{ij}(j-i) = \sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}(j-i) = -2\sum_{i=1}^{n}\sum_{j=1}^{n} i\beta_{ij}.$$

But

$$\sum_{i=1}^{n}\beta_{ij} = n+1-2p(i)$$

where $p(i)$ is the rank of the i-th object in ranking 2. This is proved in Daniels (1944), p. 131.

We therefore have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} i\beta_{ij} = \sum_{i=1}^{n} i\{n+1-2p(i)\}$$
$$= 2\left\{\tfrac{1}{4}n(n+1)^2 - \sum_{i=1}^{n} ip(i)\right\}.$$

But the Spearman rank correlation coefficient between rankings 2 and 3 which we denote by r_{s23} is

$$r_{s23} = \frac{\sum_{i=1}^{n} ip(i) - \tfrac{1}{4}n(n+1)^2}{\tfrac{1}{12}n(n^2-1)}$$

and therefore

$$\sum_{i<j} b_{ij}(j-i) = \tfrac{1}{6}n(n^2-1)r_{s23}.$$

We thus find $$\sum_1^6 c_i = \tfrac{1}{6}n(n-1)\{t_{23}+\tfrac{2}{3}(n+1)r_{s23}\}$$

and $$E({}^tR^2_{1(23)}) = \frac{2}{9n(n-1)(1-t^2_{23})}\{2(2n+5)-6t^2_{23}-4(n+1)t_{23}r_{s23}\}. \tag{11}$$

This is a somewhat surprising result since it involves both t_{23} and r_{s23}.

I have calculated the exact distribution of ${}^tR^2_{1(23)}$ for $n = 4$, (3124) as ranking 2, and (1234) as ranking 3. In this case we have $t_{23} = 0{\cdot}3333$, $r_{23} = 0{\cdot}4$ and the value of $E({}^tR^2_{1(23)})$ given by (11) is 17/36. The observed distribution of ${}^tR^2_{1(23)}$ is given in Table 4.

Table 4. *Distribution of* ${}^tR^2_{1(23)}$ *in set of* 24 *permutations*

${}^tR^2_{1(23)}$	0	0·1667	0·3333	0·5	0·6667	1·0000
Frequency	2	4	4	8	2	4

It is easy to check from this that $E({}^tR^2_{1(23)}) = \frac{17}{36}$. It will be seen that this distribution is not so smoothly spread out as that of $t_{12.3}$ when all permutations are possible, but this is to be expected since there are only 4! cases to be considered. For larger n (how large it is difficult to say exactly) it is probably reasonable to test ${}^tR^2_{1(23)}$ as if it were the square of an ordinary multiple correlation whose expectation is equal to (11). For such a correlation coefficient we have (Kendall (1947), p. 382)

$$E(R^2_{1(23\ldots p)}) = \frac{p-1}{N-1},$$

where $p-1$ is the number of independent variates and N is the size of the sample. Then $(1-R^2_{1(23)})/R^2_{1(23)}$ is distributed as Snedecor's F with $N-p$ and $p-1$ degrees of freedom. To obtain an approximation to the distribution of ${}^tR^2_{1(23)}$ we therefore put $p-1 = 2$ and define N^* by

$$E({}^tR^2_{1(23)}) = \frac{2}{N^*-1}.$$

We then test $$\frac{1-{}^tR^2_{1(23)}}{{}^tR^2_{1(23)}}$$

as Snedecor's F with N^*-2 and 2 degrees of freedom. Unfortunately the case $n = 4$ is too rough for any useful comparison with this approximation.

Multiple correlation based on Spearman's r_s can be discussed in the same way, but does not appear to have any advantages over the above method, for small n at least. However, when n increases, t becomes more arduous to calculate than r_s and coefficients based on Spearman coefficients may be easier to use. The expectation of the square of the Spearman multiple correlation coefficient can be found by the same kind of argument as that used above. The reason why, in both cases, multiple correlation coefficients are easier to handle than partial ones, is that in the former case it is only the numerator which is a random variable and not both numerator and denominator.

REFERENCES

DANIELS, H. E. (1944). *Biometrika*, **33**, 129.
DAVID, F. N. (1938). *Tables of the Correlation Coefficient*. Biometrika Office, London.
DAVID, S. T., KENDALL, M. G. & STUART, A. (1951). *Biometrika*, **38**, 131.
HOEFFDING, W. (1948). *Ann. Math. Statist.* **19**, 293.
HOTELLING, H. & PABST, M. R. (1936). *Ann. Math. Statist.* **7**, 29.
KENDALL, M. G. *et al.* (1939). *Biometrika*, **30**, 251.
KENDALL, M. G. (1942). *Biometrika*, **32**, 277.
KENDALL, M. G. (1947). *The Advanced Theory of Statistics*, **1**, 3rd ed. London: Charles Griffin and Co.
KENDALL, M. G. (1948). *Rank Correlation Methods*. London: Charles Griffin and Co.
KENDALL, M. G. (1949). *Biometrika*, **36**, 177.

AN APPLICATION OF THE DISTRIBUTION OF THE RANKING CONCORDANCE COEFFICIENT

BY A. STUART, *Division of Research Techniques, London School of Economics*

1. Pitman (1938) considered the analysis of variance test, from which the ranking concordance test emerges as a special case, in the hypothetical population formed by the permutation of sample values. The nature of this approach precludes its being used to test any hypothesis other than that of independence. I propose to examine the distribution of the concordance coefficient W in samples from a ranking population in which concordance exists.

In a recent paper, Daniels (1950) distinguished two types of population from which we may sample rankings. In his first type, we imagine the population of N individuals to be ranked from 1 to N according to each of m different qualities. A sample of n members is then drawn from it, and ranked from 1 to n for each quality in the same order as in the parent ranking. In Daniels's second type, we take a sample of m rankings from a population, say M, of rankings of 1 to n. Both types result in a sample of m rankings of 1 to n, but the generating processes and the sampling problems involved are essentially different.

I am here concerned solely with the second type of sampling: the distribution of W is considered as a problem in sampling m members from a population of M, each member of which bears the numbers 1 to n in some order. The problem is thus one of sampling from an n-variate population under the restriction that for each member the variate values are a permutation of the first n natural numbers. The results will then be expressed in terms of parameters of the population, and not, as Pitman's were, in terms of sample statistics.

2. I denote by k_1 the mean-statistic for any of the n variates. Where more than one variate is involved in the formulae, I write k_{10}, k_{100}, etc., and denote by Σ the operation of summation over the n variates, permutation of subscripts always being permitted in summation.

The coefficient of concordance may be defined as the ratio of the variance of the n means k_1 to the variance they would have if all rankings were identical, namely $\frac{1}{12}(n^2-1)$. Since the mean of the n means k_1 for a given sample is always $\frac{1}{2}(n+1)$, we have

$$W = \frac{12}{n(n^2-1)} \Sigma \left(k_1 - \frac{n+1}{2}\right)^2.$$

Since W is invariant under a change of origin, we may measure each k_1 from $\frac{1}{2}(n+1)$, and W becomes

$$W = \frac{12}{n(n^2-1)} \Sigma k_1^2. \tag{1}$$

We need only consider sampling fluctuations in Σk_1^2, which will be called V.

In deriving the moments of samples from a finite population, I use the method developed by Irwin & Kendall (1944). The finite population is itself regarded as a sample from a hypothetical infinite population, and the actual sample therefore becomes a two-stage sample from this infinite population. We denote by E_2 the operation of taking expectations over all possible samples from the finite population, and by E_1 the operation of taking expectations over all possible finite populations of the given size from the infinite population. Thus $E = E_1 E_2$, where E has its usual connotation over all possible samples from the infinite

population. Also, following Irwin & Kendall, we denote by K_r the correspondent of the k_r the rth sample k-statistic, calculated from the finite population values, and by κ_r the cumulant of the infinite population.

3. With this notation, we have

$$E_2(V) = \Sigma E_2(k_1^2) = \Sigma E_2\{(k_1^2 - K_1^2) + K_1^2\}. \tag{2}$$

Now $$E(k_1^2) = \frac{\kappa_2}{m} + \kappa_1^2, \quad E_1(K_1^2) = \frac{\kappa_2}{M} + \kappa_1^2.$$

Hence $$E(k_1^2 - K_1^2) = \kappa_2\left(\frac{1}{m} - \frac{1}{M}\right),$$

and thus, by the Irwin-Kendall principle,

$$E_2(k_1^2 - K_1^2) = K_2\left(\frac{1}{m} - \frac{1}{M}\right).$$

Thus, from (2), $$E_2(V) = \Sigma\left\{K_2\left(\frac{1}{m} - \frac{1}{M}\right) + K_1^2\right\}$$

and $$E_2(W) = \frac{12}{n(n^2-1)}\left\{\left(\frac{1}{m} - \frac{1}{M}\right)\Sigma K_2 + \Sigma K_1^2\right\}. \tag{3}$$

The somewhat unexpected appearance of terms in K_1^2 arises from the fact that the k_1's whose squares contribute to W have different parent means. The term ΣK_1^2 represents an inter-class effect, being, in fact, n times the variance of parent means.

For an infinite population, $M \to \infty$, $K_r \to \kappa_r$, and

$$E(W) = \frac{12}{n(n^2-1)}\left\{\frac{1}{m}\Sigma\kappa_2 + \Sigma\kappa_1^2\right\}. \tag{4}$$

If, on the other hand, the population is finite, and all rankings are equi-frequent in it, all the K_1's are zero, our origin for each being at $\frac{1}{2}(n+1)$, and (3) reduces to

$$E_2(W) = \frac{12}{n(n^2-1)}\left(\frac{1}{m} - \frac{1}{M}\right)\Sigma K_2. \tag{5}$$

Further, all K_2's are the same, and equal to $\frac{M}{M-1}\frac{(n^2-1)}{12}$, and hence

$$E_2(W) = \frac{M-m}{M-1}\frac{1}{m}. \tag{6}$$

This is not the result given by Pitman (1938) and Kendall (1948), namely,

$$E(W) = \frac{1}{m}.$$

They were, in fact, effectively considering sampling with replacement, and no correction for finiteness arose. Equation (6) implies that in the type of case we are considering, the mean value of Spearman's coefficient r_s in samples of two rankings from a finite population is not quite zero. Using the relation for r_s between the $\frac{1}{2}m(m-1)$ pairs of values of a set of m rankings, namely,

$$\text{average } r_s = \frac{mW-1}{m-1}$$

and putting $m = 2$, we have $$r_s = 2W - 1,$$

whence, from (6),
$$E_2(r_s) = -\frac{1}{M-1}. \tag{7}$$

Both (6) and (7), of course, reduce to the more usual results when M tends to infinity. The correction for finiteness is generally negligible.

4. It will be seen from (1) and (3) that
$$W - E_2(W) = \frac{12}{n(n^2-1)} \Sigma \left\{ k_1^2 - \left[\left(\frac{1}{m} - \frac{1}{M} \right) K_2 + K_1^2 \right] \right\}. \tag{8}$$

We require the moments of W, that is to say, the expectation of powers of the right-hand side of (8). It will readily be appreciated that the presence of terms in K_1^2 makes the moments of W much more complicated than they would otherwise be. Furthermore, in taking expectations of powers of k_1, we cannot take the mean of each k_1 to be zero, and the relations between sampling-moments and sampling cumulants for these powers will themselves also contain terms in K_1. Either of these difficulties alone would make our task laborious: together, they imply that the sampling moments of W in the general case are altogether too long and cumbrous to be of much practical use.

Henceforth, therefore, I consider only the case where the means of the population rankings are the same. This is not of course to say that we are dealing only with the case of equifrequent rankings. The hypothesis of interest here is that in the population there are no differences between the mean rankings of the n objects; it specifies homogeneity of means but not homogeneity of variances.

5. In the equal-mean case, $E_2(W)$ is given by (5).

For the variance of V we have
$$E_2(V^2) = E_2(\Sigma k_1^2)^2 = E_2\{\Sigma k_1^4 + \Sigma\Sigma k_{10}^2 k_{01}^2\}. \tag{9}$$

Now by the ordinary relation between moments and cumulants,
$$E(k_1^4) = \kappa(1^4) + 3\kappa^2(1^2)$$
$$= \frac{\kappa_4}{m^3} + \frac{3\kappa_2^2}{m^2}.$$

$E_2(k_1^4)$ is that function of the K_r whose expectation is $E(k_1^4)$. We know that
$$\left.\begin{aligned} E_1(K_4) &= \kappa_4, \\ E_1(K_2^2) &= \kappa(K_2^2) + \kappa_2^2 = \frac{\kappa_4}{M} + \frac{M+1}{M-1}\kappa_2^2. \end{aligned}\right\} \tag{10}$$

It is clear from (10) that
$$E_1\left\{ \left(\frac{M-1}{M+1} \right) \frac{3}{m^2} K_2^2 + \left[\frac{1}{m^3} - \left(\frac{M-1}{M+1} \right) \frac{3}{Mm^2} \right] K_4 \right\} = \frac{\kappa_4}{m^3} + \frac{3\kappa_2^2}{m^2}. \tag{11}$$

It follows that
$$E_2(k_1^4) = \left(\frac{M-1}{M+1} \right) \frac{3}{m^2} K_2^2 + \left[\frac{1}{m^3} - \left(\frac{M-1}{M+1} \right) \frac{3}{Mm^2} \right] K_4. \tag{12}$$

The process for deriving $E_2(k_{10}^2 k_{01}^2)$ is similar but more complicated. We start from the relation
$$\mu_{22} = \kappa_{22} + 2\kappa_{11}^2 + \kappa_{20}\kappa_{02},$$
which may be obtained by Kendall's (1940) operative process. Applying this, we have
$$E(k_{10}^2 k_{01}^2) = \kappa \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} + 2\kappa^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \kappa \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \kappa \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}. \tag{13}$$

Using Rule 10 for the sampling cumulants of k-statistics (Kendall, 1943, p. 265), we find

$$\left.\begin{aligned} \kappa\begin{pmatrix}1&1&0&0\\0&0&1&1\end{pmatrix} &= \frac{\kappa_{22}}{m^3},\\ \kappa\begin{pmatrix}1&0\\0&1\end{pmatrix} &= \frac{\kappa_{11}}{m},\\ \kappa\begin{pmatrix}1&1\\0&0\end{pmatrix} &= \frac{\kappa_{20}}{m}. \end{aligned}\right\} \tag{14}$$

Substituting (14) in (13), we obtain

$$E(k_{10}^2 k_{01}^2) = \frac{\kappa_{22}}{m^3} + \frac{2\kappa_{11}^2}{m^2} + \frac{\kappa_{20}\kappa_{02}}{m^2}. \tag{15}$$

We have that

$$\left.\begin{aligned} E_1(K_{22}) &= \kappa_{22},\\ E_1(K_{11}^2) &= \kappa(K_{11}^2) + \kappa_{11}^2 = \frac{\kappa_{22}}{M} + \frac{M}{M-1}\kappa_{11}^2 + \frac{\kappa_{20}\kappa_{02}}{M-1},\\ E_1(K_{20}K_{02}) &= \kappa(K_{20}K_{02}) + \kappa_{20}\kappa_{02} = \frac{\kappa_{22}}{M} + \frac{2\kappa_{11}^2}{M-1} + \frac{\kappa_{20}\kappa_{02}}{M-1}, \end{aligned}\right\} \tag{16}$$

where the sampling cumulants have been evaluated by Kendall's operative process.

For the right-hand side of (15) we require the coefficient of κ_{11}^2 to be twice that of $\kappa_{20}\kappa_{02}$. In (16) these coefficients are

		κ_{11}^2	$\kappa_{20}\kappa_{02}$
(*a*)	$E_1(K_{11}^2)$	$\frac{M}{M-1}$	$\frac{1}{M-1}$
(*b*)	$E_1(K_{20}K_{02})$	$\frac{2}{M-1}$	1

so we need only solve the equation

$$\frac{M}{M-1}a + \frac{2}{M-1}b = 2\left\{\frac{1}{M-1}a + b\right\},$$

which gives

$$\frac{a}{b} = 2.$$

We therefore have

$$E_1\{2K_{11}^2 + K_{20}K_{02}\} = \frac{3\kappa_{22}}{M} + 2\left(\frac{M+1}{M-1}\right)\kappa_{11}^2 + \left(\frac{M+1}{M-1}\right)\kappa_{20}\kappa_{02},$$

whence the right-hand side of (15) is equal to

$$E_1\left\{\left(\frac{M-1}{M+1}\right)\frac{1}{m^2}[2K_{11}^2 + K_{20}K_{02}] + \left[\frac{1}{m^3}\left(\frac{M-1}{M+1}\right)\frac{3}{Mm^2}K_{22}\right]\right\} \tag{17}$$

and

$$E_2(k_{10}^2 k_{01}^2) = \left(\frac{M-1}{M+1}\right)\frac{1}{m^2}[2K_{11}^2 + K_{20}K_{02}] + \left[\frac{1}{m^3} - \left(\frac{M-1}{M+1}\right)\frac{3}{Mm^2}\right]K_{22}. \tag{18}$$

6. If we now substitute the results (12) and (18) in (9), we obtain, on subtracting $\{E_2(V)\}^2$,

$$\begin{aligned} \mathrm{var}_2(V) = \Sigma\left\{\left[\frac{1}{m^3} - \frac{3}{m^2M}\left(\frac{M-1}{M+1}\right)\right]K_4 + \left[\frac{3}{m^2}\left(\frac{M-1}{M+1}\right) - \left(\frac{1}{m} - \frac{1}{M}\right)^2\right]K_2^2\right\}\\ + \Sigma\Sigma\left\{\left[\frac{1}{m^3} - \frac{3}{m^2M}\left(\frac{M-1}{M+1}\right)\right]K_{22} + \left[\frac{1}{m^2}\left(\frac{M-1}{M+1}\right) - \left(\frac{1}{m} - \frac{1}{M}\right)^2\right]K_{20}K_{02} + \frac{2}{m^2}\left(\frac{M-1}{M+1}\right)K_{11}^2\right\}. \end{aligned} \tag{19}$$

The variance of W is, of course, $\left\{\frac{12}{n(n^2-1)}\right\}^2$ times (19). If the population is infinite, (19) gives

$$\text{var}(W) = \left\{\frac{12}{mn(n^2-1)}\right\}^2 \left\{\Sigma\left(\frac{\kappa_4}{m} + 2\kappa_2^2\right) + \Sigma\Sigma\left(\frac{\kappa_{22}}{m} + 2\kappa_{11}^2\right)\right\}. \tag{20}$$

If, in addition, all rankings are equally frequent in the population, we have (Stuart, 1950)

$$\left.\begin{aligned} \kappa_2 &= \tfrac{1}{12}(n^2-1), \\ \kappa_4 &= \tfrac{1}{120}(1-n^4), \end{aligned}\right\} \tag{21}$$

while it is easy to show that

$$\left.\begin{aligned} \kappa_{11} &= -\tfrac{1}{12}(n+1), \\ \kappa_{22} &= -\tfrac{1}{360}(n+1)(n+3)(2n-1). \end{aligned}\right\} \tag{22}$$

Substituting (21) and (22) in (20), we obtain for the infinite equi-frequent population

$$\text{var}(W) = \frac{2(m-1)}{m^3(n-1)}, \tag{23}$$

which is Pitman's (1938) result.

7. It would evidently be a straightforward, although extremely laborious, task to obtain the higher moments of W for a finite population, but the expression for the variance (19) suggests that they would be too much complicated by the finite population factors to be of use. We therefore confine ourselves henceforth tc infinite populations.

For the third moment of V, we have

$$E\{V-E(V)\}^3 = E\left\{\Sigma\left(k_1^2 - \frac{\kappa_2}{m}\right)\right\}^3$$

and on expanding the bracket, substituting for the sampling moments in terms of sampling cumulants by means of the relations

$$\left.\begin{aligned} \mu_6 &= \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3, \\ \mu_{42} &= \kappa_{42} + \kappa_{40}\kappa_{02} + 8\kappa_{31}\kappa_{11} + 4\kappa_{30}\kappa_{12} + 6\kappa_{22}\kappa_{20} + 6\kappa_{21}^2 + 3\kappa_{20}^2\kappa_{02} + 12\kappa_{20}\kappa_{11}^2, \\ \mu_{222} &= \kappa_{222} + \kappa_{220}\kappa_{002} + \kappa_{202}\kappa_{020} + \kappa_{022}\kappa_{200} + 4\kappa_{211}\kappa_{011} + 4\kappa_{121}\kappa_{101} + 4\kappa_{112}\kappa_{110} + 2\kappa_{210}\kappa_{012} \\ &\quad + 2\kappa_{201}\kappa_{021} + 2\kappa_{120}\kappa_{102} + 4\kappa_{111}^2 + \kappa_{200}\kappa_{020}\kappa_{002} + 2\kappa_{200}\kappa_{011}^2 + 2\kappa_{020}\kappa_{101}^2 \\ &\quad + 2\kappa_{002}\kappa_{110}^2 + 8\kappa_{110}\kappa_{101}\kappa_{011} \end{aligned}\right\} \tag{24}$$

(μ_{42} and μ_{222} being obtained from μ_6 by Kendall's operative process), we find, on evaluating the sampling cumulants as before,

$$\begin{aligned} E\{V-E(V)\}^3 &= \Sigma\left(\frac{\kappa_6}{m^5} + \frac{12\kappa_4\kappa_2}{m^4} + \frac{10\kappa_3^2}{m^4} + \frac{8\kappa_2^3}{m^3}\right) \\ &\quad + 3\Sigma\Sigma\left(\frac{\kappa_{42}}{m^5} + \frac{8\kappa_{31}\kappa_{11}}{m^4} + \frac{4\kappa_{30}\kappa_{12}}{m^4} + \frac{4\kappa_{22}\kappa_{20}}{m^4} + \frac{6\kappa_{21}^2}{m^4} + \frac{8\kappa_{20}\kappa_{11}^2}{m^3}\right) \\ &\quad + \Sigma\Sigma\Sigma\left(\frac{\kappa_{222}}{m^5} + \frac{12\kappa_{211}\kappa_{011}}{m^4} + \frac{6\kappa_{210}\kappa_{012}}{m^4} + \frac{4\kappa_{111}^2}{m^4} + \frac{8\kappa_{110}\kappa_{101}\kappa_{011}}{m^3}\right). \end{aligned} \tag{25}$$

As might be expected, the expression for the fourth moment of V is exceedingly complicated, involving sampling moments of the k_1's like $\mu_8, \mu_{62}, \mu_{44}, \mu_{422}$, and μ_{2222}. The formulae connecting these moments with sampling cumulants would alone occupy a page of this journal.

I have therefore evaluated the fourth moment to order (m^{-4}), and omitting all details, give only the result:

$$\mu_4(V) = \frac{12}{m^4}\{5\Sigma\kappa_2^4 + \Sigma\Sigma(\kappa_{20}^2\kappa_{02}^2 + 8\kappa_{20}\kappa_{02}\kappa_{11}^2 + 6\kappa_{11}^4 + 20\kappa_{20}^2\kappa_{11}^2) + 2\Sigma\Sigma\Sigma(\kappa_{200}^2\kappa_{011}^2 + 8\kappa_{200}\kappa_{110}\kappa_{101}\kappa_{011} + 6\kappa_{110}^2\kappa_{101}^2) + \Sigma\Sigma\Sigma\Sigma(\kappa_{1100}^2\kappa_{0011}^2 + 4\kappa_{1100}\kappa_{1010}\kappa_{0101}\kappa_{0011})\}. \tag{26}$$

8. Each of the moments of V is a sum of products of cumulants of the same total order. We may, as an approximation, neglect terms below a certain order in m. Now the only terms in $\mu_3(V)$ involving cumulants of higher order than 2 are of lower order than (m^{-3}), as can be seen from (25). Similarly, (20) shows that in the variance terms down to order (m^{-2}) contain only second order cumulants, while the mean value of V is in any case dependent only on the second order cumulants. If we approximate by neglecting lower orders than those indicated, we know that if it were possible for the distributions of rankings allotted in the population for the n objects to be of the normal form, our approximations would in fact be exact expressions since all cumulants of order greater than two would then be zero. Thus if the distributions are unimodal, and not too far from normality, we may expect our approximations to be fairly good for moderate sample sizes. We find, to the above orders of approximation,

$$\left.\begin{aligned} E(V) &= \frac{1}{m}\Sigma\kappa_2, \\ \operatorname{var}(V) &= \frac{2}{m^2}\{\Sigma\kappa_2^2 + \Sigma\Sigma\kappa_{11}^2\}, \end{aligned}\right\} \tag{27}$$

and

$$\mu_3(V) = \frac{8}{m^3}\{\Sigma\kappa_2^3 + 3\Sigma\Sigma\kappa_{20}\kappa_{11}^2 + \Sigma\Sigma\Sigma\kappa_{110}\kappa_{101}\kappa_{011}\}. \tag{28}$$

We check (26) and (28) by substituting in them the values (21) and (22) for the equi-frequent population. We find

$$\mu_3(W) = \frac{8}{m^3(n-1)^2}, \tag{29}$$

$$\mu_4(W) = \frac{12(n+3)}{m^4(n-1)^3}, \tag{30}$$

which agree with Pitman's results to the same orders in m.

9. We can use the expressions (26)–(28) to test the hypothesis that the n objects have equal mean ranks $\frac{1}{2}(n+1)$ in the population, and, if this hypothesis is not discredited, to test further hypotheses about the second-order cumulants of the population, its variances and covariances. We may, for example, wish to test whether the rank allotted to one object influences the rank allotted to another, i.e. whether their covariance is zero; or we may wish to test whether the variance of one ranking is greater than that of another. In either case, we can insert hypothetical values for the characteristics under test into the moments of W, and estimate the remaining cumulants by the k-statistics of the sample. We can then fit a Pearson frequency-curve by the methods of moments, find significance points by some suitable transformation, and see whether the observed value falls outside these points. If so, the hypothesis is rejected.

We must, in all cases, first test whether the means can have been equal in the population. If this is not so, the distribution of W can be of no further practical use to us. Even in this limited case of the distribution, the computation involved is heavy enough, as will be seen in a later illustrative example.

If we wish to test only the narrower equi-frequent hypothesis, we can use Friedman's (1937) result that $m(n-1)\,W$ is distributed like χ^2 with $(n-1)$ degrees of freedom. Of course, if this is not rejected, the wider equal-mean hypothesis must also stand, while if it is overwhelmingly rejected, the latter is also likely to be rejected. In view of the fairly heavy computation involved, it is worth making the simple equi-frequent test first to avoid unnecessary computation.

10. We consider now the practical problem which gave rise to the preceding theory. A recent inquiry by the Social Research Division of the London School of Economics, which is reported by Hall & Caradog Jones (1950), was concerned with ideas of social stratification. A sample of approximately 1000 people were asked to rank 30 selected occupations in terms of 'social class' which was not defined. Effectively, the mean rank allotted to each occupation was calculated, and the problem then arose: How far were the differences between the sample mean ranks significant of real differences in the population?

It is clear that, as above, we can treat this as an ordinary case of sampling from a population of rankers. We, therefore, have at once that, for any object, the sample mean rank is an unbiased estimator of the population mean, and that its variance is given by

$$\operatorname{var} k_1 = \frac{M-m}{M-1}\frac{\kappa_2}{m}. \tag{31}$$

We have also that the covariance of any two object means is a similar function of the covariance of the ranks allotted to these objects in the population, i.e.

$$\operatorname{cov}(k_{1,0}k_{01}) = \frac{M-m}{M-1}\frac{\kappa_{11}}{m}. \tag{32}$$

Where the number of objects being ranked is as large as here, we expect the difference between two mean ranks to be distributed approximately normally by a central limit effect, with variance obtained from (31) and (32) as

$$\operatorname{var}(k_{10}-k_{01}) = \frac{M-m}{M-1}\frac{1}{m}(\kappa_{20}+\kappa_{02}-2\kappa_{11}). \tag{33}$$

For an infinite population we obtain from this, on substitution of sample estimators,

$$\operatorname{var}(k_{10}-k_{01}) = \frac{1}{m}(k_{20}+k_{02}-2k_{11}). \tag{34}$$

For $n = 2$, this reduces to the binomial test, as it should. Several comments need to be made about this test, quite apart from the emphasis on its approximate character. First, we can only test means in pairs, and $\binom{n}{2}$ may be a large number. We can avoid most of the labour in practice, for after a few obvious tests have been carried out, the eye is often enough to test widely separated means. Secondly, the terms in (34) will in general be different for each pair of objects: we cannot automatically conclude, if $|k_{100}-k_{010}|$ is significant and $|k_{100}-k_{001}| > |k_{100}-k_{010}|$, that $|k_{100}-k_{001}|$ will also be significant. Nevertheless, this is very generally true. Thirdly, as always where non-independent tests are involved, we cannot directly compound probabilities, but must rely to a certain extent on common-sense judgement.

11. We chose a non-random sample of 50 rankings from the data obtained in the original inquiry. Further, to reduce computation, only eight occupations are considered, their original rankings having been renumbered from 1 to 8. This procedure is permissible, since

we require the data for illustrative purposes only. We thus have a sample of 50 rankings of eight occupations, these being

A: Bricklayer	*E*: Commercial Traveller
B: Chartered Accountant	*F*: Elementary School-teacher
C: Chef	*G*: Farmer
D: Civil Servant	*H*: Tractor Driver

We first wish to test whether this sample could have arisen from a population in which all occupations had the same mean rank. The Friedman χ^2 test is so highly significant that we should not normally carry out the equal-mean test, but we proceed in this case as an illustration.

12. From the 50 rankings, we find the eight k_1, eight k_2 and twenty-eight k_{11} to be:

Table 1

Occupation	k_1	k_2	k_{11}							
			A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
A	7·02	0·8771	—	−1·072	−0·1304	−0·126	−0·4828	+0·0172	+0·1792	−0·2096
B	1·36	0·4392	—	—	+0·1128	−0·268	+0·2296	−0·0304	−0·4744	−0·1928
C	5·52	1·4792	—	—	—	−0·356	−0·1328	−0·7728	−0·0208	−0·1496
D	2·30	0·7857	—	—	—	—	+0·118	+0·158	−0·372	+0·076
E	5·14	1·3882	—	—	—	—	—	−0·1596	−0·7456	−0·1872
F	4·14	1·3473	—	—	—	—	—	—	−0·5456	+0·0128
G	3·04	1·6310	—	—	—	—	—	—	—	+0·0808
H	7·48	0·5812	—	—	—	—	—	—	—	—

From these values we calculate estimates of the quantities involved in equations (26)–(28) as follows:

$$\Sigma k_2 = 8{\cdot}5290, \qquad \Sigma\Sigma\Sigma k_{110}k_{101}k_{011} = 0{\cdot}8310,$$
$$\Sigma k_2^2 = 11{\cdot}2559, \qquad \Sigma\Sigma\Sigma k_{200}^2 k_{011}^2 = 39{\cdot}1475,$$
$$\Sigma k_2^3 = 15{\cdot}1754, \qquad \Sigma\Sigma\Sigma k_{200}k_{110}k_{101}k_{011} = 1{\cdot}4599,$$
$$\Sigma k_2^4 = 21{\cdot}4386, \qquad \Sigma\Sigma\Sigma k_{110}^2 k_{101}^2 = 1{\cdot}3824,$$
$$\Sigma\Sigma k_{11}^2 = 2{\cdot}6224, \qquad \Sigma\Sigma\Sigma\Sigma k_{1100}^2 k_{0011}^2 = 15{\cdot}1979,$$
$$\Sigma\Sigma k_{20}k_{11}^2 = 6{\cdot}6862, \qquad \Sigma\Sigma\Sigma\Sigma k_{1100}k_{1010}k_{0101}k_{0011} = 0{\cdot}4314.$$
$$\Sigma\Sigma k_{20}^2 k_{02}^2 = 62{\cdot}2355,$$
$$\Sigma\Sigma k_{20}k_{02}k_{11}^2 = 8{\cdot}1836,$$
$$\Sigma\Sigma k_{11}^4 = 1{\cdot}8224,$$
$$\Sigma\Sigma k_{20}^2 k_{11}^2 = 9{\cdot}9432,$$

We then find, for our estimates of the moments of V,

$$E(V) = 0{\cdot}17058,$$
$$\text{var}(V) = 0{\cdot}01110,$$
$$\mu_3(V) = 0{\cdot}00231,$$
$$\mu_4(V) = 0{\cdot}00111,$$

giving

$$\beta_1 = 3{\cdot}8950,$$
$$\beta_2 = 9{\cdot}0334.$$

The criterion κ (Elderton, 1938, chart facing p. 51) is found to be greater than unity, and

$$2\beta_2 - 3\beta_1 - 6 = 0{\cdot}3818.$$

For a Type III curve, this quantity should be zero, so bearing in mind that we have only estimated the moments, a Type III curve should give a reasonably good fit. V is therefore distributed in the form, with origin at its mean,

$$dP \propto \left(1 + \frac{V}{A}\right)^{\gamma a} e^{-\gamma V} dV.$$

The substitution $\left(1 + \frac{V}{A}\right) = x$ shows that $2\gamma A x$ is distributed like χ^2 with $2(\gamma a + 1)$ degrees of freedom.

From Elderton's formulae, we find

$$\gamma A = \gamma a + 1 = 1{\cdot}02696,$$
$$\gamma = 9{\cdot}61039,$$
$$A = 0{\cdot}10686.$$

Writing $(0{\cdot}17058 + V)$ for V to bring the origin of measurement to zero, we see that $19{\cdot}22078(0{\cdot}27744 + V)$ may be tested in the χ^2 distribution with 2·05 degrees of freedom.

The observed value of $V = \Sigma\left(k_1 - \frac{n+1}{2}\right)^2$ is 33·6416.

The resulting value of χ^2, which is over 600, is extremely significant. We conclude that the sample could not have arisen from a population with equal mean ranks.

13. Having established the heterogeneity of the means taken all together, we can use (34) to test their homogeneity in groups.

We first consider the group *BDG*, with mean ranks of 1·36, 2·30 and 3·04 in our sample. From Table 1, we find:

Objects	Difference of means	Standard error
BD	0·94	0·19
DG	0·74	0·25
BG	1·68	0·25

All three differences are significant at the 1 % level, and the group is therefore regarded as heterogeneous.

For the group *ECAH*, with mean ranks 5·14, 5·52, 7·02 and 7·48, we see at once that all non-adjacent pairs are significantly different. The adjacent pairs give the table:

Objects	Difference of means	Standard error
EC	0·38	0·25
CA	1·10	0·23
AH	0·46	0·20

EC is not significant, *CA* is highly significant, and *AH* is roughly significant at the 1 % level. Thus of the six pairs, four differences are highly significant, and the group is adjudged heterogeneous.

14. *Summary.* A test is proposed to decide whether a set of m rankings of n objects can probably have arisen from a population in which the mean rankings of the objects were equal. A second, more approximate test, can then be used to examine the mean ranks of the objects in pairs and larger groups.

I am much indebted to Prof. M. G. Kendall for criticism and advice during the preparation of this paper.

REFERENCES

DANIELS, H. E. (1950). Rank correlation and population models. *J. R. Statist. Soc.* B, **12**, 171.
ELDERTON, W. P. (1938). *Frequency Curves and Correlation*, 3rd ed. Cambridge University Press.
FRIEDMAN, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.* **32**, 675.
HALL, J. & CARADOG JONES, D. (1950). Social grading of occupations. *Brit. J. Sociology*, **1**, 31.
IRWIN, J. O. & KENDALL, M. G. (1944). Sampling moments of moments for a finite population. *Ann. Eugen., Lond.*, **12**, 138.
KENDALL, M. G. (1940). The derivation of multivariate sampling formulae from univariate formulae by symbolic operation. *Ann. Eugen., Lond.*, **10**, 392.
KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin and Co.
KENDALL, M. G. (1948). *Rank Correlation Methods*. London: Charles Griffin and Co.
PITMAN, E. J. G. (1938). The analysis of variance test. *Biometrika*, **29**, 332.
STUART, A. (1950). The cumulants of the first n natural numbers. *Biometrika*, **37**, 446.

THE EFFECT OF NON-NORMALITY ON THE POWER FUNCTION OF THE F-TEST IN THE ANALYSIS OF VARIANCE

By F. N. DAVID and N. L. JOHNSON

1. The effect of departure from normality in the distribution of the residual, or error term, on the significance levels of z (or F) used in the analysis of variance has been studied for the simple case of a one-way classification by Pearson (1931), Geary (1947) and lately by Gayen (1950). The power function of the F-test, based on the distribution of F when the null hypothesis of equal means is not true, was considered by Tang (1938). More recently Patnaik (1949) developed an approximate method of obtaining the power function of F based on a χ^2 approximation to the non-central χ^2 distribution. Both Tang and Patnaik assumed that the error terms were normally distributed with a common variance from group to group. It has seemed of interest to us to attack the problem in a more general way by studying the distribution of F when the expectations, the variances, and also the higher cumulants of the distributions of the error terms may vary from group to group. Our results cover as a special (simplified) case the interesting problem of the power function in the analysis of variance when there are unequal 'within group' standard deviations. This has been studied, for normal variation, by Quensel (1947). In this paper we give our theoretical approach and results, and some slight illustration. The main numerical results are reserved for a further publication.

2. While giving some indication in §4 of how the theory may be extended, we confine ourselves here to the one-way classification. If it is supposed that there are available s groups containing $n_1, n_2, \ldots, n_s$ observations respectively, where

$$\sum_{t=1}^{s} n_t = N,$$

then the one-way classification leads to the following identity:

$$\sum_{t=1}^{s} n_t(\bar{z}_{t.} - \bar{z}_{..})^2 + \sum_{t=1}^{s} \sum_{i=1}^{n_t} (z_{ti} - \bar{z}_{t.})^2 = \sum_{t=1}^{s} \sum_{i=1}^{n_t} (z_{ti} - \bar{z}_{..})^2.$$

It is usual in analysis of variance tests to assume that all the variables z_{ti} have the same normal distribution. We shall generalize this by assuming that the n_t variables

$$z_{ti}, \quad (t = 1, 2, \ldots, s) \quad (i = 1, 2, \ldots, n_t)$$

each have the same distribution which we may specify by as many cumulants $\kappa_{1t}, \kappa_{2t}, \ldots, \kappa_{rt}, \ldots$ as desired, but that the values of these cumulants may be different for different groups. It will be noticed we are assuming that if the κ_{1t} are not equal, (i.e. if the null hypothesis does not hold), systematic rather than random effects are present so that the κ_{1t} are treated as parameters. For brevity we shall write

$$S_1 = \sum_{t=1}^{s} n_t(\bar{z}_{t.} - \bar{z}_{..})^2, \quad S_2 = \sum_{t=1}^{s} \sum_{i=1}^{n_t} (z_{ti} - \bar{z}_{t.})^2$$

and

$$f_1 = s - 1, \quad f_2 = N - s$$

for the degrees of freedom of S_1 and S_2 respectively.

Ideally we should investigate the ratio

$$F = (S_1/f_1)/(S_2/f_2)$$

but it is difficult to obtain its distribution for other than the normal case, and its moments can usually only be obtained as approximations. In the course of our work we shall indicate how approximations to the moments of the distribution of z (or F) may be obtained, but for the main part we shall rewrite the ratio in order to work with exact quantities.

3. Let F_α be the value of F at the normal probability level α. If the usual null normal hypothesis is true then

$$P\{(S_1/f_1)/(S_2/f_2) \geqslant F_\alpha\} = \alpha$$

or

$$P\left\{\left(S_1 - \frac{f_1 F_\alpha}{f_2} S_2\right) \geqslant 0\right\} = \alpha$$

or writing

$$-\frac{f_1}{f_2} F_\alpha = a$$

where a is a constant for given f_1, f_2 and α, we have

$$P\{S_1 + aS_2 \geqslant 0\} = \alpha.$$

This suggests that we should consider the distribution of $S_1 + aS_2$ in the non-normal case. The actual distribution of this quantity is difficult if not impossible to obtain, but the exact moments of the distribution are a matter of simple, if laborious, algebra. We have found these moments and our procedure will be to fit a Pearson (or some other suitable) curve to these moments. We shall have therefore, for each f_1 and f_2, to fit a new curve for each separate value chosen for α, but, provided a suitable system of curves is chosen for graduation purposes, the labour involved is not heavy.

4. We might perhaps note here that the algebraic results obtained in this present problem of the one-way classification, and the method used, may be applied to other more complicated analysis of variance problems. For example, if we consider a second-order hierarchical classification we may write

$$\sum_{i=1}^{s} n_{i.}(\bar{z}_{i\ .} - \bar{z}_{...})^2 + \sum_{i=1}^{s} \sum_{i=1}^{m_i} n_{ii}(\bar{z}_{ii.} - \bar{z}_{i..})^2 + \sum_{i=1}^{s} \sum_{i=1}^{m_i} \sum_{j=1}^{n_{ii}} (z_{iij} - \bar{z}_{ii.})^2 = \sum_{i=1}^{s} \sum_{i=1}^{m_i} \sum_{j=1}^{n_{ii}} (z_{iij} - \bar{z}_{...})^2$$

where

$$n_{i.} = \sum_{i=1}^{m_i} n_{ii}.$$

The second expression, the 'between subgroups within groups' sum of squares can be considered as a sum of statistics like S_1, independent, and therefore with additive cumulants. The third expression, the 'residual' sum of squares, can be considered as a sum of independent statistics like S_2. Similarly the cross-products can be obtained from the cross-products of S_1 and S_2 although care will be necessary to ensure the enumeration of all the terms involved.

5. It will only be necessary to outline the derivation of the moments of

$$S = S_1 + aS_2$$

briefly. We require

$$\begin{aligned}\mu(S^r) &= \mathscr{E}[S - \mathscr{E}(S)]^r = \mathscr{E}\{[S_1 - \mathscr{E}(S_1)] + a[S_2 - \mathscr{E}(S_2)]\}^r \\ &= \mathscr{E}[S_1 - \mathscr{E}(S_1)]^r + {}^nC_1 a\mathscr{E}[S_1 - \mathscr{E}(S_1)]^{r-1}[S_2 - \mathscr{E}(S_2)] + \ldots + a^r\mathscr{E}[S_2 - \mathscr{E}(S_2)]^r.\end{aligned}$$

For condensation we shall use the customary notation

$$\mu(S_1^b S_2^c) = \mathscr{E}\{[S_1 - \mathscr{E}(S_1)]^b [S_2 - \mathscr{E}(S_2)]^c\}$$

and use $\kappa(S_1^b S_2^c)$ for the corresponding cumulant. It will only be necessary to use cumulants for $b+c \geqslant 4$, but for these cases the expressions for the cumulants are much shorter than those for the corresponding moments. Partly because of the weight of the algebra and partly because the individual values may be used in other ways, as we indicated in the previous section, we have evaluated the moments of S_1, S_2 and their joint moments separately. We begin by a transformation of variables. We have assumed that

$$\mathscr{E}(z_{ti}) = \kappa_{1t} = \mathscr{E}(\bar{z}_{t.}), \quad \mathscr{E}(\bar{z}_{..}) = \frac{1}{N}\sum_{t=1}^{s}\sum_{i=1}^{n_t}\mathscr{E}(z_{ti}) = \frac{1}{N}\sum_{t=1}^{s} n_t\kappa_{1t} = \kappa_1 \quad \text{(say)}$$

where N is the total number of observations. Let

$$x_{ti} = z_{ti} - \kappa_{1t}, \quad \kappa_{1t} - \kappa_1 = C_t, \quad \frac{1}{N}\sum_{t=1}^{s} n_t(\kappa_{1t}-\kappa_1)^2 = \sigma_c^2, \quad \sum_{t=1}^{s} n_t C_t = 0.$$

Substitution into S_1 gives

$$S_1 = \sum_{t=1}^{s} n_t\left(\frac{1}{n_t}\sum_{i=1}^{n_t} x_{ti}\right)^2 - \frac{1}{N}\left(\sum_{t=1}^{s}\sum_{i=1}^{n_t} x_{ti}\right)^2 + N\sigma_c^2 + 2\sum_{t=1}^{s}\sum_{i=1}^{n_t} C_t x_{ti}$$

or defining a k-statistic for each group, i.e. writing

$$k_{1t} = \frac{1}{n_t}\sum_{i=1}^{n_t} x_{ti}$$

we have

$$S_1 = \sum_{t=1}^{s} n_t k_{1t}^2 - \frac{1}{N}\left(\sum_{t=1}^{s} n_t k_{1t}\right)^2 + 2\sum_{t=1}^{s} n_t k_{1t} C_t + N\sigma_c^2.$$

and

$$\mathscr{E}(S_1) = \sum_{t=1}^{s} \kappa_{2t}\left(1-\frac{n_t}{N}\right) + N\sigma_c^2.$$

The first quantity with which we have to work is therefore

$$S_1 - \mathscr{E}(S_1) = \sum_{t=1}^{s} n_t k_{1t}^2 - \frac{1}{N}\left(\sum_{t=1}^{s} n_t k_{1t}\right)^2 + 2\sum_{t=1}^{s} n_t k_{1t} C_t - \sum_{t=1}^{s} \kappa_{2t}\left(1-\frac{n_t}{N}\right).$$

Similarly

$$S_2 = \sum_{t=1}^{s}\sum_{i=1}^{n_t} (x_{ti}-\bar{x}_{t.})^2,$$

whence if we define a second k-statistic

$$k_{2t} = \frac{1}{n_t - 1}\sum_{i=1}^{n_t} (x_{ti}-\bar{x}_{t.})^2$$

we have

$$S_2 = \sum_{t=1}^{s} (n_t-1)\,k_{2t}, \quad \mathscr{E}(S_2) = \sum_{t=1}^{s} (n_t-1)\,\kappa_{2t}, \quad S_2 - \mathscr{E}(S_2) = \sum_{t=1}^{s} (n_t-1)\,(k_{2t}-\kappa_{2t}).$$

6. Moments and product moments are obtained by multiplying out the appropriate powers of the two quantities S_1 and S_2 and taking expectations, a procedure which is made easy by the use of David & Kendall's tables of symmetric functions (1949). We give the results below:

$$\mathscr{E}(S_1) = \sum_{t=1}^{s} \kappa_{2t}\left(1-\frac{n_t}{N}\right) + N\sigma_c^2, \quad \mathscr{E}(S_2) = \sum_{t=1}^{s} (n_t-1)\,\kappa_{2t},$$

$$\mathscr{E}(S) = \sum_{t=1}^{s} \kappa_{2t}\left[1-a+n_t\left(a-\frac{1}{N}\right)\right] + N\sigma_c^2,$$

$$\mu(S_1^2) = \kappa(S_1^2) = \sum_{t=1}^{s} \frac{\kappa_{4t}}{n_t}\left(1-\frac{n_t}{N}\right)^2 + 2\sum_{t=1}^{s} \kappa_{2t}^2\left(1-\frac{2n_t}{N}\right) + \frac{2}{N^2}\left(\sum_{t=1}^{s} n_t\kappa_{2t}\right)^2 + 4\sum_{t=1}^{s} \kappa_{2t} C_t\left(1-\frac{n_t}{N}\right)$$
$$+ 4\sum_{t=1}^{s} n_t\kappa_{2t} C_t^2,$$

$$\mu(S_1 S_2) = \sum_{i=1}^{s} \frac{n_i - 1}{n_i}\left(1 - \frac{n_i}{N}\right)\kappa_{4i} + 2\sum_{i=1}^{s}(n_i - 1)C_i\kappa_{3i},$$

$$\mu(S_2^2) = \sum_{i=1}^{s}(n_i - 1)^2\left(\frac{\kappa_{4i}}{n_i} + \frac{2\kappa_{2i}^2}{n_i - 1}\right),$$

$$\mu_2(S) = \sum_{i=1}^{s}\frac{\kappa_{4i}}{n_i}\left[1 - a + n_i\left(a - \frac{1}{N}\right)\right]^2 + 2\sum_{i=1}^{s}\kappa_{2i}^2\left[1 - a^2 + n_i\left(a^2 - \frac{2}{N}\right)\right] + \frac{2}{N^2}\left(\sum_{i=1}^{s} n_i\kappa_{2i}\right)^2$$
$$+ 4\sum_{i=1}^{s}\kappa_{3i}C_i\left[1 - a + n_i\left(a - \frac{1}{N}\right)\right] + 4\sum_{i=1}^{s} n_i\kappa_{2i}C_i^2.$$

It is of interest to note that if we put $C_i = 0, \kappa_{ri} = \kappa_r, r = 1, 2, \ldots$, we get

$$\mu(S_1^2) = \kappa_4\left(\sum_{i=1}^{s}\frac{1}{n_i} - \frac{(2s-1)}{N}\right) + 2\kappa_2^2(s-1),$$

$$\mu(S_1 S_2) = \kappa_4\left(s - 1 + \frac{s}{N} - \sum_{i=1}^{s}\frac{1}{n_i}\right),$$

$$\mu(S_2^2) = \kappa_4\left(\sum_{i=1}^{s}\frac{1}{n_i} + N - 2s\right) + 2\kappa_2^2(N - s),$$

results which were first given by Pearson (1931),

$$\mu(S_1^3) = \sum_{i=1}^{s}\frac{\kappa_{6i}}{n_i^2}\left(1 - \frac{n_i}{N}\right)^3 + 12\sum_{i=1}^{s}\frac{\kappa_{4i}\kappa_{2i}}{n_i}\left(1 - \frac{2n_i}{N}\right)\left(1 - \frac{n_i}{N}\right) + 2\sum_{i=1}^{s}\frac{\kappa_{2i}^3}{n_i}\left(5 - \frac{12n_i}{N} + \frac{9n_i^2}{N^2}\right)$$
$$+ 8\sum_{i=1}^{s}\kappa_{2i}^3\left(1 - \frac{3n_i}{N}\right) + \frac{12}{N^2}\left(\sum_{i=1}^{s} n_i\kappa_{2i}\right)\left[\sum_{i=1}^{s}\kappa_{4i}\left(1 - \frac{n_i}{N}\right)\right] - \frac{8}{N^3}\left(\sum_{i=1}^{s} n_i\kappa_{2i}\right)^3$$
$$+ \frac{24}{N^2}\left(\sum_{i=1}^{s} n_i\kappa_{2i}\right)\left(\sum_{i=1}^{s} n_i\kappa_{2i}^2\right) - \frac{2}{N}\left[3\left(\sum_{i=1}^{s}\kappa_{2i}\right)^2 - \frac{6}{N}\left(\sum_{i=1}^{s}\kappa_{3i}\right)\left(\sum_{i=1}^{s} n_i\kappa_{3i}\right)\right.$$
$$\left. + \frac{5}{N^2}\left(\sum_{i=1}^{s} n_i\kappa_{3i}\right)^2\right] + 6\sum_{i=1}^{s} C_i\frac{\kappa_{5i}}{n_i}\left(1 - \frac{n_i}{N}\right)^2 + 24\sum_{i=1}^{s} C_i\kappa_{3i}\kappa_{2i}\left(2 - \frac{3n_i}{N}\right) + 8\sum_{i=1}^{s} n_i\kappa_{3i}C_i^3$$
$$+ \frac{24}{N^2}\left(\sum_{i=1}^{s} n_i\kappa_{2i}\right)\left(\sum_{i=1}^{s} n_i C_i\kappa_{3i}\right) - \frac{24}{N}\left(\sum_{i=1}^{s} n_i C_i\kappa_{2i}\right)\left[\sum_{i=1}^{s}\kappa_{3i}\left(1 - \frac{n_i}{N}\right)\right]$$
$$+ 12\sum_{i=1}^{s} C_i^2\kappa_{4i}\left(1 - \frac{n_i}{N}\right) + 24\sum_{i=1}^{s} n_i C_i^2\kappa_{2i}^2 - \frac{24}{N}\left(\sum_{i=1}^{s} n_i C_i\kappa_{2i}\right)^2,$$

$$\mu(S_1^2 S_2) = \sum_{i=1}^{s}\frac{\kappa_{6i}}{n_i^2}\left(1 - \frac{n_i}{N}\right)^2(n_i - 1) + 4\sum_{i=1}^{s}(n_i - 1)\frac{\kappa_{4i}\kappa_{2i}}{n_i}\left(1 - \frac{2n_i}{N}\right) + \frac{4}{N^2}\left(\sum_{i=1}^{s}(n_i - 1)\kappa_{4i}\right)\left(\sum_{i=1}^{s} n_i\kappa_{2i}\right)$$
$$+ 4\sum_{i=1}^{s}\frac{(n_i - 1)}{n_i}\left(1 - \frac{n_i}{N}\right)\kappa_{3i}^2 - \frac{4}{N}\left(\sum_{i=1}^{s}(n_i - 1)\kappa_{3i}\right)\left[\sum_{i=1}^{s}\kappa_{3i}\left(1 - \frac{n_i}{N}\right)\right]$$
$$+ 4\sum_{i=1}^{s}(n_i - 1)C_i\frac{\kappa_{5i}}{n_i}\left(1 - \frac{n_i}{N}\right) + 8\sum_{i=1}^{s}(n_i - 1)C_i\kappa_{3i}\kappa_{2i} - \frac{8}{N}\left(\sum_{i=1}^{s}(n_i - 1)\kappa_{3i}\right)\left(\sum_{i=1}^{s} n_i C_i\kappa_{2i}\right)$$
$$+ 4\sum_{i=1}^{s}(n_i - 1)C_i^2\kappa_{4i},$$

$$\mu(S_1 S_2^2) = \sum_{i=1}^{s}\frac{(n_i - 1)^2}{n_i^2}\kappa_{6i}\left(1 - \frac{n_i}{N}\right) + 4\sum_{i=1}^{s}\frac{(n_i - 1)}{n_i}\kappa_{4i}\kappa_{2i}\left(1 - \frac{n_i}{N}\right) + 4\sum_{i=1}^{s}\frac{(n_i - 1)}{n_i}\kappa_{3i}^2\left(1 - \frac{n_i}{N}\right)$$
$$+ 2\sum_{i=1}^{s}\frac{(n_i - 1)^2}{n_i}\kappa_{3i}^2 - \frac{2}{N}\left(\sum_{i=1}^{s}(n_i - 1)\kappa_{3i}\right)^2 + 2\sum_{i=1}^{s}\frac{(n_i - 1)^2}{n_i}C_i\kappa_{5i} + 8\sum_{i=1}^{s}(n_i - 1)C_i\kappa_{3i}\kappa_{2i}$$

$$\mu(S_2^3) = \sum_{i=1}^{s}\frac{(n_i - 1)^3}{n_i^2}\kappa_{6i} + 12\sum_{i=1}^{s}\frac{(n_i - 1)^2}{n_i}\kappa_{4i}\kappa_{2i} + 4\sum_{i=1}^{s}\frac{(n_i - 1)(n_i - 2)}{n_i}\kappa_{3i}^2 + 8\sum_{i=1}^{s}(n_i - 1)\kappa_{2i}^3.$$

From the above we find the third moment of S to be

$$\mu_3(S) = \sum_{i=1}^{s} \frac{\kappa_{3i}}{n_i^2}\left[1-a+n_i\left(a-\frac{1}{N}\right)\right]^3 + 12\sum_{i=1}^{s}\frac{\kappa_{4i}\kappa_{2i}}{n_i}\left[1-a+n_i\left(a-\frac{1}{N}\right)\right]\left[1-a^2+n_i\left(a^2-\frac{2}{N}\right)\right]$$
$$+\sum_{i=1}^{s}\frac{\kappa_{3i}^2}{n_i}\left\{6\left[1-a+n_i\left(a-\frac{1}{N}\right)\right]^2+(4-12a^2+8a^3)-12n_i\left(\frac{1}{N}-a^2-\frac{a^2}{N}+a^3\right)\right.$$
$$\left.+4n_i^2\left(\frac{3}{N^2}-\frac{3a^2}{N}+a^3\right)\right\}+8\sum_{i=1}^{s}\kappa_{2i}^3\left[1-a^3+n_i\left(a^3-\frac{3}{N}\right)\right]$$
$$-\frac{8}{N^3}\left(\sum_{i=1}^{s}n_i\kappa_{2i}\right)^3+\frac{24}{N^2}\left(\sum_{i=1}^{s}n_i\kappa_{2i}\right)\left(\sum_{i=1}^{s}n_i\kappa_{3i}^2\right)+8\sum_{i=1}^{s}n_i\kappa_{3i}C_i^3$$
$$+\frac{12}{N^2}\left(\sum_{i=1}^{s}n_i\kappa_{2i}\right)\left\{\sum_{i=1}^{s}\kappa_{4i}\left[1-a+n_i\left(a-\frac{1}{N}\right)\right]\right\}-\frac{4}{N^3}\left(\sum_{i=1}^{s}n_i\kappa_{3i}\right)^2$$
$$-\frac{6}{N}\left\{\sum_{i=1}^{s}\kappa_{3i}\left[1-a+n_i\left(a-\frac{1}{N}\right)\right]\right\}^2+6\sum_{i=1}^{s}C_i\frac{\kappa_{3i}}{n_i}\left[1-a+n_i\left(a-\frac{1}{N}\right)\right]^2$$
$$+24\sum_{i=1}^{s}C_i\kappa_{3i}\kappa_{2i}\left(2+a+a^2-\frac{3n_i}{N}\right)-\frac{24}{N}\left(\sum_{i=1}^{s}n_iC_i\kappa_{2i}\right)\left\{\sum_{i=1}^{s}\kappa_{3i}\left[1-a+n_i\left(a-\frac{1}{N}\right)\right]\right\}$$
$$+\frac{24}{N^2}\left(\sum_{i=1}^{s}n_i\kappa_{2i}\right)\left(\sum_{i=1}^{s}n_iC_i\kappa_{3i}\right)+12\sum_{i=1}^{s}C_i^2\kappa_{4i}\left[1-a+n_i\left(a-\frac{1}{N}\right)\right]$$
$$+24\sum_{i=1}^{s}n_iC_i^2\kappa_{2i}^2-\frac{24}{N}\left(\sum_{i=1}^{s}n_iC_i\kappa_{2i}\right)^2.$$

Continuing the procedure outlined we find that

$$\kappa(S_1^4) = \mu(S_1^4)-3\mu^2(S_1^2)$$
$$= \Sigma\frac{\kappa_{5i}}{n_i^3}\left(1-\frac{n_i}{N}\right)^4+\frac{24}{N^2}(\Sigma n_i\kappa_{2i})\left[\Sigma\frac{\kappa_{5i}}{n_i}\left(1-\frac{n_i}{N}\right)^2\right]+24\,\Sigma\frac{\kappa_{5i}\kappa_{2i}}{n_i^2}\left(1-\frac{n_i}{N}\right)^2\left(1-\frac{2n_i}{N}\right)$$
$$+8\,\Sigma\frac{\kappa_{5i}\kappa_{3i}}{n_i^2}\left(1-\frac{n_i}{N}\right)\left(7-\frac{18n_i}{N}+\frac{15n_i^2}{N^2}\right)-\frac{24}{N}(\Sigma\kappa_{3i})\left[\Sigma\frac{\kappa_{5i}}{n_i}\left(1-\frac{n_i}{N}\right)^2\right]$$
$$+\frac{8}{N^2}(\Sigma n_i\kappa_{3i})\left[\Sigma\frac{\kappa_{5i}}{n_i}\left(1-\frac{n_i}{N}\right)\left(3-\frac{7n_i}{N}\right)\right]+32\,\Sigma\frac{\kappa_{4i}^2}{n_i^2}\left(1-\frac{n_i}{N}\right)^3-\frac{32}{N}\Sigma\frac{\kappa_{4i}^2}{n_i}\left(1-\frac{n_i}{N}\right)^2$$
$$+\frac{8}{N^2}\Sigma\kappa_{4i}^2\left(1-\frac{2n_i}{N}\right)+\frac{24}{N^2}\left[\Sigma\left(1-\frac{n_i}{N}\right)\kappa_{4i}\right]^2+\frac{8}{N^4}(\Sigma n_i\kappa_{4i})^2$$
$$-\frac{48}{N^3}(\Sigma n_i\kappa_{2i})^2\left[\Sigma\kappa_{4i}\left(2-\frac{3n_i}{N}\right)\right]+\frac{96}{N^2}(\Sigma n_i\kappa_{2i})\left[\Sigma\kappa_{4i}\kappa_{2i}\left(3-\frac{4n_i}{N}\right)\right]$$
$$+48\Sigma\frac{\kappa_{4i}\kappa_{2i}^2}{n_i}\left(3-\frac{12n_i}{N}+\frac{10n_i^2}{N^2}\right)+\frac{96}{N^2}(\Sigma n_i\kappa_{2i}^2)\left[\Sigma\kappa_{4i}\left(1-\frac{n_i}{N}\right)\right]$$
$$+48\Sigma\frac{\kappa_{3i}^2\kappa_{2i}}{n_i}\left(5-\frac{16n_i}{N}+\frac{15n_i^2}{N^2}\right)+\frac{96}{N^2}(\Sigma n_i\kappa_{2i})\left[\Sigma\kappa_{3i}^2\left(2-\frac{3n_i}{N}\right)\right]$$
$$+\frac{48}{N^2}(\Sigma n_i\kappa_{2i})\left((\Sigma\kappa_{3i})^2-\frac{4}{N}(\Sigma n_i\kappa_{2i})(\Sigma\kappa_{3i})+\frac{5}{N^2}(\Sigma n_i\kappa_{3i})^2\right)$$
$$-\frac{192}{N}(\Sigma\kappa_{3i}\kappa_{2i})\left[\Sigma\kappa_{3i}\left(1-\frac{n_i}{N}\right)\right]+\frac{96}{N^2}(\Sigma n_i\kappa_{3i}\kappa_{2i})\left[\Sigma\kappa_{3i}\left(3-\frac{5n_i}{N}\right)\right]+48\Sigma\kappa_{3i}^4$$
$$-\frac{192}{N}\Sigma n_i\kappa_{3i}^4+\frac{192}{N^2}(\Sigma n_i\kappa_{3i}^2)(\Sigma n_i\kappa_{2i})-\frac{192}{N^3}(\Sigma n_i\kappa_{2i}^2)(\Sigma n_i\kappa_{2i})^2+\frac{96}{N^2}(\Sigma n_i\kappa_{2i}^2)^2$$

$$+\frac{48}{N^4}(\Sigma n_i\kappa_{2i})^4+8\Sigma C_i\frac{\kappa_{7i}}{n_i^3}\left(1-\frac{n_i}{N}\right)^3-\frac{96}{N^2}(\Sigma n_i\kappa_{2i})\left[\Sigma C_i\kappa_{5i}\left(1-\frac{n_i}{N}\right)\right]$$

$$+48\Sigma C_i\frac{\kappa_{5i}\kappa_{2i}}{n_i}\left(1-\frac{n_i}{N}\right)\left(3-\frac{5n_i}{N}\right)-\frac{48}{N}(\Sigma n_iC_i\kappa_{2i})\left[\Sigma\frac{\kappa_{5i}}{n_i}\left(1-\frac{n_i}{N}\right)^2\right]$$

$$+\frac{96}{N^2}(\Sigma n_iC_i\kappa_{3i})\left[\Sigma\kappa_{4i}\left(1-\frac{n_i}{N}\right)\right]+32\left[\Sigma C_i^2\frac{\kappa_{4i}\kappa_{3i}}{n_i}\left(8-\frac{21n_i}{N}+\frac{15n_i^2}{N^2}\right)\right]$$

$$-\frac{96}{N}(\Sigma C_i\kappa_{4i})\left[\Sigma\kappa_{3i}\left(1-\frac{n_i}{N}\right)\right]+\frac{32}{N^2}(\Sigma n_iC_i\kappa_{4i})\left[\Sigma\kappa_{3i}\left(3-\frac{5n_i}{N}\right)\right]$$

$$-\frac{192}{N^3}(\Sigma n_i\kappa_{2i})^2(\Sigma n_iC_i\kappa_{3i})+\frac{576}{N^2}(\Sigma n_i\kappa_{2i})(\Sigma n_iC_i\kappa_{3i}\kappa_{2i})+\frac{192}{N^2}(\Sigma n_i\kappa_{2i}^2)(\Sigma n_iC_i\kappa_{3i})$$

$$+576\Sigma\kappa_{3i}\kappa_{2i}^2C_i\left(1-\frac{2n_i}{N}\right)+\frac{192}{N^2}(\Sigma n_i\kappa_{2i})(\Sigma n_iC_i\kappa_{2i})\left[\Sigma\kappa_{3i}\left(1-\frac{2n_i}{N}\right)\right]$$

$$-\frac{192}{N}(\Sigma n_iC_i\kappa_{2i})\left[\Sigma\kappa_{3i}\kappa_{2i}\left(2-\frac{3n_i}{N}\right)\right]-\frac{192}{N}(\Sigma n_iC_i\kappa_{2i}^2)\left[\Sigma\kappa_{3i}\left(1-\frac{n_i}{N}\right)\right]$$

$$+24\Sigma C_i^2\frac{\kappa_{6i}}{n_i}\left(1-\frac{n_i}{N}\right)^2+\frac{96}{N^2}(\Sigma n_iC_i^2\kappa_{4i})(\Sigma n_i\kappa_{2i})+96\Sigma C_i^2\kappa_{4i}\kappa_{2i}\left(3-\frac{4n_i}{N}\right)$$

$$-\frac{192}{N}(\Sigma n_iC_i\kappa_{2i})\left[\Sigma C_i\kappa_{4i}\left(1-\frac{n_i}{N}\right)\right]+96\Sigma C_i^2\kappa_{3i}^2\left(2-\frac{3n_i}{N}\right)$$

$$-\frac{96}{N}(\Sigma n_iC_i^2\kappa_{3i})\left[\Sigma\kappa_{3i}\left(1-\frac{n_i}{N}\right)\right]+192\Sigma n_iC_i^2\kappa_{2i}^3+\frac{192}{N^2}(\Sigma n_i\kappa_{2i})(\Sigma n_iC_i\kappa_{2i})^2$$

$$-\frac{384}{N}(\Sigma n_iC_i\kappa_{2i}^2)(\Sigma n_iC_i\kappa_{2i})+32\Sigma C_i^3\kappa_{5i}\left(1-\frac{n_i}{N}\right)+192\Sigma n_iC_i^2\kappa_{3i}\kappa_{2i}$$

$$-\frac{192}{N}(\Sigma n_iC_i\kappa_{2i})(\Sigma n_iC_i^2\kappa_{3i})+16\Sigma n_iC_i^4\kappa_{4i},$$

$$\kappa(S_1^3S_2)=\mu(S_1^3S_2)-3\kappa(S_1S_2)\kappa(S_1^2)$$

$$=\Sigma\frac{(n_i-1)}{n_i^3}\kappa_{6i}\left(1-\frac{n_i}{N}\right)^3+12\Sigma\frac{(n_i-1)}{n_i^2}\kappa_{6i}\kappa_{2i}\left(1-\frac{n_i}{N}\right)\left(1-\frac{2n_i}{N}\right)$$

$$+\frac{12}{N^2}(\Sigma n_i\kappa_{2i})\left[\Sigma\frac{(n_i-1)}{n_i}\kappa_{6i}\left(1-\frac{n_i}{N}\right)\right]+\Sigma\frac{(n_i-1)}{n_i^2}\kappa_{5i}\kappa_{3i}\left(26-\frac{60n_i}{N}+\frac{42n_i^2}{N^2}\right)$$

$$-\frac{6}{N}[\Sigma(n_i-1)\kappa_{3i}]\left[\Sigma\frac{\kappa_{5i}}{n_i}\left(1-\frac{n_i}{N}\right)^2\right]+\frac{4}{N^2}[\Sigma(n_i-1)\kappa_{5i}]\left[\Sigma\kappa_{3i}\left(3-\frac{5n_i}{N}\right)\right]$$

$$-\frac{12}{N}\left(\Sigma\frac{n_i-1}{n_i}\kappa_{5i}\right)\left[\Sigma\kappa_{3i}\left(1-\frac{n_i}{N}\right)\right]+12\Sigma\frac{(n_i-1)}{n_i^2}\kappa_{4i}^2\left(1-\frac{n_i}{N}\right)\left(1-\frac{2n_i}{N}\right)$$

$$+\frac{12}{N^2}[\Sigma(n_i-1)\kappa_{4i}]\left[\Sigma\kappa_{4i}\left(1-\frac{n_i}{N}\right)\right]+24\Sigma\kappa_{4i}\kappa_{2i}^2\frac{n_i-1}{n_i}\left(1-\frac{3n_i}{N}\right)$$

$$+\frac{24}{N^2}(\Sigma n_i\kappa_{2i}^2)[\Sigma(n_i-1)\kappa_{4i}]-\frac{24}{N^3}(\Sigma n_i\kappa_{2i})^2[\Sigma(n_i-1)\kappa_{4i}]$$

$$+\frac{48}{N^2}(\Sigma n_i\kappa_{2i})[\Sigma(n_i-1)\kappa_{4i}\kappa_{2i}]+24\Sigma\kappa_{3i}^2\kappa_{2i}\frac{n_i-1}{n_i}\left(2-\frac{3n_i}{N}\right)$$

$$-\frac{24}{N}[\Sigma(n_i-1)\kappa_{3i}\kappa_{2i}]\left[\Sigma\kappa_{3i}\left(1-\frac{n_i}{N}\right)\right]+\frac{24}{N^2}(\Sigma n_i\kappa_{2i})[\Sigma(n_i-1)\kappa_{3i}]\left[\Sigma\kappa_{3i}\left(1-\frac{2n_i}{N}\right)\right]$$

$$-\frac{24}{N}[\Sigma(n_i-1)\kappa_{3i}]\left[\Sigma\kappa_{3i}\kappa_{2i}\left(2-\frac{3n_i}{N}\right)\right]+\frac{24}{N^2}[\Sigma(n_i-1)\kappa_{3i}^2](\Sigma n_i\kappa_{2i})$$

$$+6\Sigma\frac{n_i-1}{n_i^2}C_i\kappa_{7i}\left(1-\frac{n_i}{N}\right)^2+24\Sigma\frac{n_i-1}{n_i}C_i\kappa_{5i}\kappa_{2i}\left(2-\frac{3n_i}{N}\right)$$
$$-\frac{24}{N}(\Sigma n_iC_i\kappa_{2i})\left[\Sigma\frac{n_i-1}{n_i}\kappa_{5i}\left(1-\frac{n_i}{N}\right)\right]+\frac{24}{N^2}[\Sigma(n_i-1)C_i\kappa_{5i}](\Sigma n_i\kappa_{2i})$$
$$+24\Sigma\frac{(n_i-1)}{n_i}C_i\kappa_{4i}\kappa_{3i}\left(3-\frac{4n_i}{N}\right)-\frac{24}{N}[\Sigma(n_i-1)C_i\kappa_{4i}]\left[\Sigma\kappa_{3i}\left(1-\frac{n_i}{N}\right)\right]$$
$$+\frac{12}{N}[\Sigma(n_i-1)\kappa_{4i}]\left[\Sigma\kappa_{3i}C_i\left(1-\frac{n_i}{N}\right)\right]-\frac{24}{N}[\Sigma(n_i-1)\kappa_{3i}]\left[\Sigma C_i\kappa_{4i}\left(1-\frac{n_i}{N}\right)\right]$$
$$-\frac{12}{N}[\Sigma(n_i-1)\kappa_{4i}]\left[\Sigma C_i\kappa_{3i}\left(1-\frac{3n_i}{N}\right)\right]+48\Sigma(n_i-1)C_i\kappa_{3i}\kappa_{2i}^2$$
$$+\frac{48}{N^2}(\Sigma n_i\kappa_{2i})(\Sigma n_iC_i\kappa_{2i})[\Sigma(n_i-1)\kappa_{3i}]-\frac{48}{N}(\Sigma n_iC_i\kappa_{2i}^2)[\Sigma(n_i-1)\kappa_{3i}]$$
$$-\frac{48}{N}[\Sigma(n_i-1)\kappa_{3i}\kappa_{2i}](\Sigma n_iC_i\kappa_{2i})$$
$$+12\Sigma\frac{n_i-1}{n_i}C_i^2\kappa_{6i}\left(1-\frac{n_i}{N}\right)+48\Sigma(n_i-1)C_i^2\kappa_{4i}\kappa_{2i}-\frac{48}{N}[\Sigma(n_i-1)C_i\kappa_{4i}](\Sigma n_iC_i\kappa_{2i})$$
$$+24\Sigma(n_i-1)C_i^2\kappa_{3i}^2-\frac{24}{N}[\Sigma(n_i-1)\kappa_{3i}](\Sigma n_iC_i^2\kappa_{3i})$$
$$+8\Sigma(n_i-1)C_i^2\kappa_{5i},$$

$$\kappa(S_1^2S_2^2)=\mu(S_1^2S_2^2)-2\kappa^2(S_1S_2)-\kappa(S_1^2)\kappa(S_2^2)$$
$$=\Sigma\frac{(n_i-1)^2}{n_i^3}\kappa_{8i}\left(1-\frac{n_i}{N}\right)^2+4\Sigma\kappa_{6i}\kappa_{2i}(n_i-1)\left(\frac{1}{n_i}+\frac{1}{N^2}-\frac{2}{N}\right)+\frac{4}{N^2}\left(\Sigma\frac{(n_i-1)^2}{n_i}\kappa_{6i}\right)(\Sigma n_i\kappa_{2i})$$
$$+4\Sigma\kappa_{5i}\kappa_{3i}\frac{(n_i-1)}{n_i^2}\left(1-\frac{n_i}{N}\right)\left(1-\frac{4n_i}{N}+3n_i\right)-\frac{4}{N}\left(\Sigma\frac{(n_i-1)^2}{n_i}\kappa_{5i}\right)\left[\Sigma\kappa_{3i}\left(1-\frac{n_i}{N}\right)\right]$$
$$-\frac{8}{N}[\Sigma(n_i-1)\kappa_{3i}]\left[\Sigma\kappa_{5i}\left(1-\frac{n_i}{N}\right)\left(\frac{n_i-1}{n_i}\right)\right]$$
$$+4\Sigma\kappa_{4i}^2\frac{(n_i-1)}{n_i}\left[3\left(1-\frac{n_i}{N}\right)^2+(n_i-1)\left(1-\frac{2n_i}{N}\right)\right]+\frac{4}{N^2}[\Sigma(n_i-1)\kappa_{4i}]^2$$
$$+16\Sigma\kappa_{4i}\kappa_{2i}^2\frac{n_i-1}{n_i}\left(1-\frac{2n_i}{N}\right)+\frac{16}{N^2}(\Sigma n_i\kappa_{2i})[\Sigma\kappa_{4i}\kappa_{2i}(n_i-1)]+16\Sigma\kappa_{3i}^2\kappa_{2i}\frac{n_i-1}{n_i}\left(2-\frac{3n_i}{N}\right)$$
$$+8\Sigma\kappa_{3i}^2\kappa_{2i}\frac{(n_i-1)^2}{n_i}-\frac{16}{N}[\Sigma(n_i-1)\kappa_{3i}\kappa_{2i}]\left[\Sigma n_i\kappa_{3i}\left(1-\frac{1}{N}\right)\right]$$
$$+\frac{16}{N^2}(\Sigma n_i\kappa_{2i})[\Sigma(n_i-1)\kappa_{3i}^2]+\frac{8}{N^2}(\Sigma n_i\kappa_{2i})[\Sigma(n_i-1)\kappa_{3i}]^2$$
$$+4\Sigma\kappa_{7i}C_i\frac{(n_i-1)^2}{n_i^2}\left(1-\frac{n_i}{N}\right)+16\Sigma\kappa_{5i}\kappa_{2i}C_i\frac{n_i-1}{n_i}\left(1-\frac{n_i}{N}\right)+8\Sigma\kappa_{5i}\kappa_{2i}C_i\frac{(n_i-1)^2}{n_i}$$
$$-\frac{8}{N}\left(\Sigma\frac{(n_i-1)^2}{n_i}\kappa_{5i}\right)(\Sigma n_iC_i\kappa_{2i})+48\Sigma\kappa_{4i}\kappa_{3i}C_i\frac{n_i-1}{n_i}\left(1-\frac{n_i}{N}\right)+16\Sigma\kappa_{4i}\kappa_{3i}C_i\frac{(n_i-1)^2}{n_i}$$
$$-\frac{16}{N}[\Sigma(n_i-1)C_i\kappa_{4i}][\Sigma(n_i-1)\kappa_{3i}]+32\Sigma\kappa_{3i}\kappa_{2i}^2C_i(n_i-1)$$
$$-\frac{32}{N}[\Sigma(n_i-1)\kappa_{3i}\kappa_{2i}](\Sigma n_iC_i\kappa_{2i})+4\Sigma C_i^2\kappa_{6i}\frac{(n_i-1)^2}{n_i}$$
$$+16\Sigma C_i^2(n_i-1)\kappa_{4i}\kappa_{2i}+16\Sigma C_i^2\kappa_{3i}^2(n_i-1),$$

$$\kappa(S_1 S_2^3) = \mu(S_1 S_2^3) - 3\kappa(S_1 S_2)\,\kappa(S_2^2)$$

$$= \sum\kappa_{8i}\frac{(n_i-1)^3}{n_i^3}\left(1-\frac{n_i}{N}\right) + 12\sum\kappa_{6i}\kappa_{2i}\frac{(n_i-1)^2}{n_i^3}\left(1-\frac{n_i}{N}\right) + 6\sum\kappa_{5i}\kappa_{3i}\frac{(n_i-1)^3}{n_i^3}$$

$$+ 8\sum\kappa_{5i}\kappa_{3i}\frac{(4n_i-5)(n_i-1)}{n_i^3}\left(1-\frac{n_i}{N}\right) - \frac{6}{N}\left(\sum\frac{(n_i-1)^2}{n_i}\kappa_{5i}\right)[\sum(n_i-1)\kappa_{3i}]$$

$$+ 4\sum\kappa_{4i}^2\frac{(5n_i-7)(n_i-1)}{n_i^3}\left(1-\frac{n_i}{N}\right) + 24\sum\frac{\kappa_{4i}\kappa_{2i}^2}{n_i}\left(1-\frac{n_i}{N}\right) + 24\sum\kappa_{3i}^2\kappa_{2i}\frac{(n_i-1)^2}{n_i}$$

$$+ 48\sum\kappa_{3i}^2\kappa_{2i}\frac{(n_i-1)}{n_i}\left(1-\frac{n_i}{N}\right) - \frac{24}{N}[\sum\kappa_{3i}\kappa_{2i}(n_i-1)][\sum\kappa_{3i}(n_i-1)]$$

$$+ 2\sum\frac{(n_i-1)^3}{n_i^3}C_i\kappa_{7i} + 24\sum\frac{(n_i-1)^2}{n_i}C_i\kappa_{5i}\kappa_{2i} + 8\sum\kappa_{4i}\kappa_{3i}C_i\frac{(n_i-1)(5n_i-7)}{n_i}$$

$$+ 48\sum C_i\kappa_{3i}\kappa_{2i}^2(n_i-1),$$

$$\kappa(S_2^4) = \mu(S_2^4) - 3\kappa^2(S_2^2)$$

$$= \sum\kappa_{8i}\frac{(n_i-1)^4}{n_i^3} + 24\sum\kappa_{6i}\kappa_{2i}\frac{(n_i-1)^3}{n_i^3} + 32\sum\kappa_{5i}\kappa_{3i}\frac{(n_i-2)(n_i-1)^2}{n_i^3}$$

$$+ 8\sum\kappa_{4i}^2\frac{(n_i-1)}{n_i^3}(4n_i^2-9n_i+6) + 144\sum\kappa_{4i}\kappa_{2i}^2\frac{(n_i-1)^2}{n_i}$$

$$+ 96\sum\kappa_{3i}^2\kappa_{2i}\frac{(n_i-2)(n_i-1)}{n_i} + 48\sum\kappa_{2i}^4(n_i-1).$$

The expression for the fourth moment of S may be reached by converting the cumulants given above into moments and by combining these fourth order moments as previously described. The expression, as will be realized, is one of great length, and for this reason we do not give it here. It is considered more important to give the individual expressions as these will be of use in investigations other than that with which we are concerned.

7. The moments as given above are quite general, the only assumption which is made being that the cumulants of the distribution of x in each group do exist and are the same for each individual of the group. We now proceed to make certain simplifying assumptions which will enable different types of distributions to be studied which are useful in statistical theory. We first assume

$$\kappa_{ri} = \kappa_r \quad \text{for} \quad i = 1, 2, \ldots, s, \quad r \geqslant 2,$$

or in other words we assume that each x has the same distribution for each group. Under this simplifying assumption we find the second-order moments as given in the preceding section and, remembering $\sum n_i C_i = 0$, we have also

$$\mu(S_1^3) = \mathscr{E}[S_1 - \mathscr{E}(S_1)]^3 = \kappa_6\left(\sum\frac{1}{n_i^2} - \frac{3}{N}\sum\frac{1}{n_i} + \frac{3s-1}{N^2}\right) + 12\kappa_4\kappa_2\left(\sum\frac{1}{n_i} - \frac{2s}{N} + \frac{1}{N}\right)$$

$$+ 2\kappa_3^2\left(\sum\frac{5}{n_i} - \frac{3s^2}{N} - \frac{6s}{N} + \frac{4}{N}\right) + 8\kappa_3^2(s-1) + 6\kappa_5\left(\sum\frac{C_i}{n_i} - \frac{2}{N}\sum C_i\right)$$

$$+ 48\kappa_3\kappa_2\sum C_i + 12\kappa_4\sum C_i^2\left(1-\frac{n_i}{N}\right) + 24\kappa_3^2\sum n_i C_i^2 + 8\kappa_3\sum n_i C_i^3,$$

$$\mu(S_1^2 S_2) = \kappa_6\left[-\sum\frac{1}{n_i^2} + \sum\frac{1}{n_i}\left(1+\frac{2}{N}\right) - \frac{s}{N}\left(2+\frac{1}{N}\right) + \frac{1}{N}\right] + 4\kappa_4\kappa_2\left[s\left(1+\frac{1}{N}\right) - 1 - \sum\frac{1}{n_i}\right]$$

$$+ 4\kappa_3^2\left(\frac{s^2}{N} - \sum\frac{1}{n_i}\right) + 4\kappa_5\left[\sum C_i\left(1+\frac{1}{N}\right) - \sum\frac{C_i}{n_i}\right] - 8\kappa_3\kappa_2\sum C_i + 4\kappa_4\sum(n_i-1)C_i^3,$$

$$\mu(S_1S_2^2) = \kappa_6\left[\Sigma\frac{1}{n_i^2}-\left(2+\frac{1}{N}\right)\Sigma\frac{1}{n_i}+s\left(1+\frac{2}{N}\right)-1\right]+4\kappa_4\kappa_2\left[s\left(1+\frac{1}{N}\right)-1-\Sigma\frac{1}{n_i}\right]$$
$$+2\kappa_3^2\left[2s\left(1+\frac{1}{N}\right)-\frac{s^2}{N}-\Sigma\frac{1}{n_i}-2\right]+2\kappa_5\left(-2\Sigma C_i+\Sigma\frac{C_i}{n_i}\right)-8\kappa_3\kappa_2\Sigma C_i,$$

$$\mu(S_2^3) = \kappa_6\left(-\Sigma\frac{1}{n_i^2}+3\Sigma\frac{1}{n_i}-3s+N\right)+12\kappa_4\kappa_2\left(\Sigma\frac{1}{n_i}-2s+N\right)+4\kappa_3^2\left(\Sigma\frac{2}{n_i}-3s+N\right)$$
$$+8\kappa_2^3(N-s).$$

When $n_i = n$, that is when the numbers are the same in all groups, then

$$\mu(S_1^3) = \kappa_6\frac{(s-1)^3}{n^2s^2}+12\kappa_4\kappa_2\frac{(s-1)^2}{ns}+4\kappa_3^2\frac{(s-1)(s-2)}{ns}+8\kappa_2^3(s-1)$$
$$+12\kappa_4\frac{s-1}{s}\Sigma C_i^2+24\kappa_2^2n\Sigma C_i^2+8\kappa_2n\Sigma C_i^3,$$

$$\mu(S_1^2S_2) = \kappa_6\frac{(s-1)^2(n-1)}{n^2s}+4\kappa_4\kappa_2\frac{(n-1)(s-1)}{n}+4\kappa_4(n-1)\Sigma C_i^2,$$

$$\mu(S_1S_2^2) = \kappa_6\frac{(n-1)^2(s-1)}{n^2}+4\kappa_4\kappa_2\frac{(n-1)(s-1)}{n}+4\kappa_3^2\frac{(n-1)(s-1)}{n},$$

$$\mu(S_2^3) = \kappa_6\frac{s(n-1)^3}{n^2}+12\kappa_4\kappa_2\frac{s(n-1)^2}{n}+4\kappa_3^2\frac{s(n-1)(n-2)}{n}+8\kappa_2^3s(n-1).$$

8. The cumulants and product cumulants of the fourth order follow similarly. We omit $\kappa(S_1^4)$ because of its great length. The other four are of reasonable proportions and we have

$$\kappa(S_1^3S_2) = \kappa_8\Sigma\frac{n_i-1}{n_i^3}\left(1-\frac{n_i}{N}\right)^3+12\kappa_6\kappa_2\Sigma\frac{n_i-1}{n_i}\left(1-\frac{n_i}{N}\right)\left(\frac{1}{n_i}-\frac{1}{N}\right)$$
$$+\kappa_5\kappa_3\left[26\Sigma\frac{1}{n_i^2}(n_i-1)-\frac{60}{N}\left(s-\Sigma\frac{1}{n_i}\right)+\frac{42}{N^2}(N-s)+\frac{4}{N^2}(N-s)(3s-5)\right.$$
$$\left.-\frac{12}{N}(s-1)\left(s-\Sigma\frac{1}{n_i}\right)-\frac{6}{N}(N-s)\Sigma\frac{1}{n_i}\left(1-\frac{n_i}{N}\right)^2\right]$$
$$+12\kappa_4^2\left[\Sigma\frac{n_i-1}{n_i^2}\left(1-\frac{n_i}{N}\right)\left(1-\frac{2n_i}{N}\right)+\frac{1}{N^2}(N-s)(s-1)\right]$$
$$+24\kappa_4\kappa_3^2\left[\Sigma\frac{n_i-1}{n_i}\left(1-\frac{3n_i}{N}\right)+\frac{2}{N}(N-s)\right]$$
$$+24\kappa_3^2\kappa_2\left[\Sigma\frac{n_i-1}{n_i}\left(2-\frac{3n_i}{N}\right)-\frac{1}{N}(N-s)(2s-3)\right]$$
$$+6\kappa_7\left[\Sigma C_i\frac{n_i-1}{n_i^2}\left(1-\frac{n_i}{N}\right)^2\right]+24\kappa_5\kappa_2\Sigma C_i\left[\frac{n_i-1}{n_i}\left(2-\frac{3n_i}{N}\right)-\frac{1}{N}\right]$$
$$+24\kappa_4\kappa_3\left[\Sigma C_i\frac{n_i-1}{n_i}\left(3-\frac{4n_i}{N}\right)\right]-48\kappa_3\kappa_2^2\Sigma C_i+12\kappa_6\Sigma\frac{n_i-1}{n_i}C_i^2\left(1-\frac{n_i}{N}\right)$$
$$+48\kappa_4\kappa_2\Sigma(n_i-1)C_i^2+24\kappa_3^2\left(\frac{s}{N}\Sigma n_iC_i^2-\Sigma C_i^2\right)+8\kappa_5\Sigma(n_i-1)C_i^3,$$

$$\kappa(S_1^2 S_2^2) = \kappa_8 \Sigma \frac{(n_t-1)^2}{n_t^3}\left(1-\frac{n_t}{N}\right)^2 + 4\kappa_6\kappa_2\left[\Sigma\frac{n_t-1}{n_t} + \frac{N-s}{N}\left(\frac{1}{N}-2\right) + \frac{1}{N}\Sigma\frac{(n_t-1)^2}{n_t}\right]$$
$$+ 4\kappa_5\kappa_3\left[\Sigma\frac{n_t-1}{n_t^2}\left(1-\frac{n_t}{N}\right)\left(1-\frac{4n_t}{N}+3n_t\right)\right.$$
$$\left. -\frac{s-1}{N}\Sigma\frac{(n_t-1)^2}{n_t} - \frac{2(N-s)}{N}\Sigma\frac{n_t-1}{n_t}\left(1-\frac{n_t}{N}\right)\right]$$
$$+ 4\kappa_4^2\left\{\Sigma\frac{n_t-1}{n_t^2}\left[3\left(1-\frac{n_t}{N}\right)^2 + (n_t-1)\left(1-\frac{2n_t}{N}\right)\right] + \left(\frac{N-s}{N}\right)^2\right\}$$
$$+ 16\kappa_4\kappa_2^2\left[\Sigma\frac{n_t-1}{n_t}\left(1-\frac{2n_t}{N}\right) + \frac{N-s}{N}\right]$$
$$+ 8\kappa_3^2\kappa_2\left[2\Sigma\frac{n_t-1}{n_t}\left(2-\frac{3n_t}{N}\right) + \Sigma\frac{(n_t-1)^2}{n_t} - \frac{(N-s)(N+s-4)}{N}\right]$$
$$+ 4\kappa_7\left[\Sigma C_t\frac{(n_t-1)^2}{n_t^2}\left(1-\frac{n_t}{N}\right)\right] + 8\kappa_5\kappa_2\left[2\Sigma C_t\frac{n_t-1}{n_t}\left(1-\frac{n_t}{N}\right) + \Sigma C_t\frac{(n_t-1)^2}{n_t}\right]$$
$$+ 16\kappa_4\kappa_3\left[3\Sigma C_t\frac{n_t-1}{n_t}\left(1-\frac{n_t}{N}\right) + \Sigma C_t\frac{(n_t-1)^2}{n_t} + \frac{N-s}{N}\Sigma C_t\right] - 32\kappa_3\kappa_2^2\Sigma C_t$$
$$+ 4\kappa_6\Sigma C_t^2\frac{(n_t-1)^2}{n_t} + 16\kappa_4\kappa_2\Sigma C_t^2(n_t-1) + 16\kappa_3^2\Sigma C_t^2(n_t-1),$$

$$\kappa(S_1 S_2^3) = \kappa_8\Sigma\frac{(n_t-1)^3}{n_t^3}\left(1-\frac{n_t}{N}\right) + 12\kappa_6\kappa_2\Sigma\frac{(n_t-1)^2}{n_t^2}\left(1-\frac{n_t}{N}\right) + 2\kappa_5\kappa_3$$
$$\times\left[3\Sigma\frac{(n_t-1)^3}{n_t^2} + 4\Sigma\frac{(4n_t-5)(n_t-1)}{n_t}\left(1-\frac{n_t}{N}\right) - \frac{3}{N}\left(\Sigma\frac{(n_t-1)^2}{n_t}\right)[\Sigma(n_t-1)\kappa_{5t}\kappa_{2t}]\right]$$
$$+ 4\kappa_4^2\left[\Sigma\frac{(5n_t-7)(n_t-1)}{n_t^2}\left(1-\frac{n_t}{N}\right)\right] + 24\kappa_4\kappa_2^2\Sigma\frac{1}{n_t}\left(1-\frac{n_t}{N}\right)$$
$$+ 24\kappa_3^2\kappa_2\left[\Sigma\frac{(n_t-1)^2}{n_t} + 2\Sigma\frac{n_t-1}{n_t}\left(1-\frac{n_t}{N}\right) - \frac{(N-s)^2}{N}\right]$$
$$+ 2\kappa_7\Sigma\frac{(n_t-1)^3}{n_t^2}C_t + 24\kappa_5\kappa_2\Sigma\frac{(n_t-1)^2}{n_t}C_t + 8\kappa_4\kappa_3\Sigma C_t\frac{(n_t-1)(5n_t-7)}{n_t}$$
$$+ 48\kappa_3\kappa_2^2\Sigma C_t(n_t-1),$$

$$\kappa(S_2^4) = \kappa_8\Sigma\frac{(n_t-1)^4}{n_t^3} + 24\kappa_6\kappa_2\Sigma\frac{(n_t-1)^3}{n_t^2} + 32\kappa_5\kappa_3\Sigma\frac{(n_t-2)(n_t-1)^2}{n_t^2}$$
$$+ 8\kappa_4^2\Sigma\frac{n_t-1}{n_t^2}(4n_t^2-9n_t+6) + 144\kappa_4\kappa_2^2\Sigma\frac{(n_t-1)^2}{n_t}$$
$$+ 96\kappa_3^2\kappa_2\Sigma\frac{(n_t-2)(n_t-1)}{n_t} + 48\kappa_2^4(N-s).$$

9. These expressions simplify still further if we take $n_t = n$ for $t = 1, 2, \ldots, s$, and we have

$$\kappa(S_1^4) = \kappa_8\frac{(s-1)^4}{(ns)^3} + 24\kappa_6\kappa_2\frac{(s-1)^3}{n^2s^2} + 32\kappa_5\kappa_3\frac{(s-1)^2(s-2)}{n^2s^2} + 8\kappa_4^2\frac{(s-1)}{n^2s^2}[4(s-1)^2-(s-2)]$$
$$+ 144\kappa_4\kappa_2^2\frac{(s-1)^2}{ns} + 96\kappa_3^2\kappa_2\frac{(s-1)(s-2)}{ns} + 48\kappa_2^4(s-1)$$
$$+ (\Sigma C_t^2)\left(24\kappa_6\frac{(s-1)^2}{ns^2} + 288\kappa_4\kappa_2\frac{s-1}{s} + 96\kappa_3^2\frac{(s-2)}{s} + 192n\kappa_2^3\right)$$
$$+ (\Sigma C_t^3)\left(32\kappa_5\frac{(s-1)}{s} + 192\kappa_3\kappa_2 n\right) + 16n\kappa_4\Sigma C_t^4,$$

$$\kappa(S_1^3 S_2) = \kappa_8 \frac{(n-1)(s-1)^3}{n^3 s^2} + 12\kappa_6\kappa_2 \frac{(n-1)(s-1)^2}{n^2 s} + 8\kappa_5\kappa_3 \frac{(s-1)(s-2)(n-1)}{n^2 s}$$
$$+ \frac{12\kappa_4^2}{n^2 s}(n-1)(s-1)^2 + 24\kappa_4\kappa_2^2 \frac{(n-1)(s-1)}{n}$$
$$+ 12\kappa_6 \frac{(n-1)(s-1)}{ns} \Sigma C_t^2 + 48\kappa_4\kappa_2(n-1)\Sigma C_t^2 + 8\kappa_5(n-1)\Sigma C_t^3,$$

$$\kappa(S_1^2 S_2^2) = \kappa_8 \frac{(n-1)^2(s-1)^2}{n^3 s} + 4\kappa_6\kappa_2 \frac{(n-1)(s-1)(N-1)}{n^2 s} + 16\kappa_5\kappa_3 \frac{(n-1)(s-1)^2}{n^2 s}$$
$$+ 4\kappa_4^2 \frac{(n-1)(s-1)}{n^2 s}[(N-1)+2(s-1)] + 16\kappa_4\kappa_2^2 \frac{(n-1)(s-1)}{n}$$
$$+ 16\kappa_3^2\kappa_2 \frac{(n-1)(s-1)}{n} + 4\kappa_6 \frac{(n-1)^2}{n} \Sigma C_t^2 + 16\kappa_4\kappa_2(n-1)\Sigma C_t^2 + 16\kappa_3^2(n-1)\Sigma C_t^3,$$

$$\kappa(S_1 S_2^3) = \kappa_8 \frac{(n-1)^3(s-1)}{n^3} + 12\kappa_6\kappa_2 \frac{(n-1)^2(s-1)}{n^2} + 32\kappa_5\kappa_3 \frac{(n-1)^2(s-1)}{n^2}$$
$$- 8\kappa_5\kappa_3 \frac{(n-1)(s-1)}{n^2} + 20\kappa_4^2 \frac{(n-1)^2(s-1)}{n^2} - 8\kappa_4^2 \frac{(n-1)(s-1)}{n^2}$$
$$+ 24\kappa_4\kappa_2^2 \frac{(s-1)(n-1)}{n} + 48\kappa_3^2\kappa_2 \frac{(n-1)(s-1)}{n},$$

$$\kappa(S_2^4) = \kappa_8 \frac{s(n-1)^4}{n^3} + 24\kappa_6\kappa_2 \frac{s(n-1)^3}{n^2} + 32\kappa_5\kappa_3 \frac{s(n-2)(n-1)^2}{n^2}$$
$$+ 8\kappa_4^2 \frac{s(n-1)}{n^2}[4(n-1)^2-(n-2)] + 144\kappa_4\kappa_2^2 \frac{s(n-1)^2}{n} + 96\kappa_3^2\kappa_2 \frac{s(n-2)(n-1)}{n}$$
$$+ 48\kappa_2^4 s(n-1).$$

10. A knowledge of the first four cumulants of a frequency distribution does not necessarily imply a knowledge of the distribution. We propose to estimate

$$P\{(S_1 + aS_2) > 0\}$$

by finding a frequency curve which will have the correct first four moments of $S_1 + aS_2$. Since we do not know, however, how good such an approximate procedure will be, we confine our calculations at first to such cases where the answer is known by other methods so that a direct check will be possible. The only arithmetical answers that are known to us are where the populations generating the groups are normal and have the same variance, so that a further simplification of the moments given above is necessary, by writing

$$\kappa_{rt} = 0 \quad r \geqslant 3, \quad t = 1, 2, \ldots, s$$

and

$$\kappa_{2t} = \kappa_2 \qquad t = 1, 2, \ldots, s.$$

Making the necessary simplifications and substituting numerical values it is found that the appropriate Pearson curves lie in the Type IV area. We therefore for simplicity of calculation consider two other systems of frequency curves, the Gram-Charlier Type A and the S_U system derived by Johnson (1949). While the former may be expected to give good results in most cases, the latter system is used as a check on whether such divergencies from exact values as are found are due to method or are due to the negative frequencies implied by a Gram-Charlier expansion being of non-negligible importance.

A numerical comparison with certain of the values quoted by Patnaik (1949) Table 7, and derived from Tang's (1938) basic tables is given below in Table 1. The symbol, λ, used by Patnaik is defined by

$$\lambda = \sum_{i=1}^{s} n_i C_i^2 / \kappa_2.$$

In our case the number of groups is 4 and the number of observations $\sum n_i = 24$, corresponding to Patnaik's case $f_1 = 3, f_2 = 20$.

These probabilities are generally slightly closer to the exact values than those obtained by using the approximation suggested by Patnaik and not quoted here. The figures show that for the calculation of power curves the technique employed for estimating probabilities is likely to be adequate.

Table 1. *Comparison of probabilities estimated from frequency curves with the exact values*

Significance level α	λ	$P\{(S_1 + aS_2) > 0\}$		
		Gram-Charlier	Johnson (S_U)	Exact
0·05	0	0·0523	0·0512	0·0500
	4	0·290	0·305	0·300
	16	0·877	0·873*	0·874
0·01	0	0·0095	0·0097	0·0100
	4	0·112	0·114	0·113
	16	0·657	0·654	0·653

* Log-normal curve fitted in this case.

11. The moments given above are so general that to list all the simplifications which might be made would require much space. We shall give two illustrations here of the way in which they may be used to attack problems concerning the single classification in the analysis of variance, and for our first illustration we consider the effect of unequal variances. It is assumed that there are s normal populations with unequal means and variances, and that the numbers drawn from the populations to form the groups are not necessarily equal. All cross-product cumulants vanish and we have

$$\mathscr{E}(S_1) = \sum_{i=1}^{s} \kappa_{2i}\left(1 - \frac{n_i}{N}\right) + N\sigma_c^2; \quad \mathscr{E}(S_2) = \sum_{i=1}^{s} (n_i - 1)\kappa_{2i},$$

$$\kappa(S_1^2) = 2\sum \kappa_{2i}^2\left(1 - \frac{2n_i}{N}\right) + \frac{2}{N^2}(\sum n_i \kappa_{2i})^2 + 4\sum n_i \kappa_{2i} C_i^2,$$

$$\kappa(S_2^2) = 2\sum (n_i - 1)\kappa_{2i}^2,$$

$$\kappa(S_1^3) = 8\sum \kappa_{2i}^3\left(1 - \frac{3n_i}{N}\right) - \frac{8}{N^3}(\sum n_i \kappa_{2i})^3 + \frac{24}{N^2}(\sum n_i \kappa_{2i})(\sum n_i \kappa_{2i}^2)$$

$$+ 24\sum n_i C_i^2 \kappa_{2i}^2 - \frac{24}{N}(\sum n_i C_i \kappa_{2i})^2,$$

$$\kappa(S_2^3) = 8\sum (n_i - 1)\kappa_{2i}^3,$$

$$\kappa(S_1^4) = 48\Sigma\kappa_{2l}^4\left(1-\frac{4n_l}{N}\right)+\frac{48}{N^4}(\Sigma n_l\kappa_{2l})^4+\frac{96}{N^2}(\Sigma n_l\kappa_{2l}^2)\left(\Sigma n_l\kappa_{2l}^2-\frac{2}{N}(\Sigma n_l\kappa_{2l})^2\right)$$
$$+\frac{192}{N^2}(\Sigma n_l\kappa_{2l}^3)(\Sigma n_l\kappa_{2l})+192\Sigma n_l C_l^2\kappa_{2l}^3$$
$$+\frac{192}{N}(\Sigma n_l C_l\kappa_{2l})\left[\left(\frac{1}{N}\Sigma n_l\kappa_{2l}\right)(\Sigma n_l C_l\kappa_{2l})-2\Sigma n_l C_l\kappa_{2l}^2\right],$$

$$\kappa(S_2^4) = 48\Sigma(n_l-1)\kappa_{2l}^4.$$

When the means of the groups are assumed equal all terms containing C_l disappear; in our brief arithmetical investigation we have assumed this to be so. Four groups each containing six observations were taken, and six combinations of values for $\kappa_{21}, \kappa_{22}, \kappa_{23}, \kappa_{24}$ were considered. The effect of inequalities in these variances on the nominal significance levels is shown in Table 2, the Gram-Charlier system being used for purposes of estimation.

Table 2. *Effect of unequal variances on the nominal significance levels of the F-test*

κ_{21}	κ_{22}	κ_{23}	κ_{24}	Nominal level 0·05	Nominal level 0·01
1	1	1	1	0·052	0·010
1	1	1	2	0·057	0·013
1	1	2	2	0·057	0·013
1	2	2	2	0·055	0·011
1	1	1	3	0·063	0·021
1	1	2	3	0·060	0·015
1	1	3	3	0·063	0·016

The first line of the table has been reproduced from Table 1 so that the accuracy of the method employed may be borne in mind. It will be noted, as might be expected, that the situation most likely to lead to wrong judgements of significance will be when one of the group variances is very much larger than all the others.

12. Gayen (1950) has discussed the effect of non-normality on the F-test when the parent population generating the samples is of Gram-Charlier type. We have assumed nothing about the shape of the parent populations, except that the cumulants exist, and that these cumulants may be different for each population generating each group, and thus it would be possible with the aid of the moments and product-moments which we have given earlier, to discuss approximations to the distribution of the F ratio to order $1/N^3$ in a very general way. We outline as a second illustration how this might be done for the z ratio, our choice of z rather than F being based on the reason that previous work by David (1949) showed that the numerical coefficients of the expansions were smaller for z than for F. Thus

$$2z = \log\{(S_1/f_1)/(S_2/f_2)\} = \log\left(1+\frac{S_1-\mathscr{E}(S_1)}{\mathscr{E}(S_1)}\right)-\log\left(1+\frac{S_2-\mathscr{E}(S_2)}{\mathscr{E}(S_2)}\right)+\log\frac{\mathscr{E}(S_1)}{f_1}-\log\frac{\mathscr{E}(S_2)}{f_2}.$$

As previously let $\mathscr{E}[S_1-\mathscr{E}(S_1)]^r[S_2-\mathscr{E}(S_2)]^r = \mu(S_1^rS_2^r)$

and further write $M(S_1^rS_2^r) = \mu(S_1^rS_2^r)/[\mathscr{E}(S_1)]^r[\mathscr{E}(S_2)]^r$

Expanding the first two logarithms on the right-hand side, a procedure which will be valid only for S and n large, and taking expectations we have

$$2\mathscr{E}(z)-\log\{[\mathscr{E}(S_1)/f_1]/[\mathscr{E}(S_2)/f_2]\} = -\tfrac{1}{2}[M(S_1^2)-M(S_2^2)] + \tfrac{1}{3}[M(S_1^3)-M(S_2^3)]-\tfrac{1}{4}[M(S_1^4)-M(S_2^4)]+\ldots.$$

If we decide to work to order $1/N^3$ there will be contributions from terms of the fifth and sixth orders for

$$\mu(S^6)=\kappa(S^6)+15\kappa(S^4)\kappa(S^2)+10\kappa^2(S^3)+15\kappa^3(S^2)$$
$$\mu(S^5)=\kappa(S^5)+10\kappa(S^2)\kappa(S^3)$$

and $\kappa(S^2)\kappa(S^3)$ and $\kappa^3(S^2)$ will both be of order $1/N^3$.
The second moment may be obtained by squaring the original expression for z, whence

$$\{2z-\log[\mathscr{E}(S_1)/f_1]/[\mathscr{E}(S_2)/f_2]\}^2 = M(S_1^2)-2M(S_1S_2)+M(S_2^2)-M(S_1^3)+M(S_1^2S_2) + M(S_1S_2^2)-M(S_2^3)+\tfrac{11}{12}M(S_1^4)-\tfrac{2}{3}M(S_1^3S_2) - \tfrac{1}{2}M(S_1^2S_2^2)-\tfrac{2}{3}M(S_1S_2^3)+\tfrac{11}{12}M(S_2^4)-\ldots.$$

The same remarks hold good about the order of the expansions. Similar expressions may be found for the third and fourth moments.

13. Particularly simple expressions result from the procedure of the previous section if it is assumed that the generating populations have the same non-normal distribution for each group, the same mean, if the numbers in the groups are the same, and if we work to order f^{-2} only. Writing

$$\gamma_r=\frac{\kappa_{r+2}}{\kappa_2^{\frac{1}{2}r+1}},\quad f_1=s-1,\quad f_2=s(n-1)$$

we shall have

$$\mathscr{E}(S_1)=\kappa_2 f_1,\quad \mathscr{E}(S_2)=\kappa_2 f_2$$

and

$$M(S_1^2)=\frac{\gamma_2}{N}+\frac{2}{f_1},\quad M(S_1S_2)=\frac{\gamma_2}{N},\quad M(S_2^2)=\frac{\gamma_2}{N}+\frac{2}{f_2},$$

$$M(S_1^3)=\frac{\gamma_4}{N^2}+\frac{12\gamma_2}{N}\frac{1}{f_1}+\frac{4\gamma_1^2}{N}\frac{f_1-1}{f_1^2}+\frac{8}{f_1^2},$$

$$M(S_1^2S_2)=\frac{\gamma_4}{N^2}+\frac{4\gamma_2}{N}\frac{1}{f_1},$$

$$M(S_1S_2^2)=\frac{\gamma_4}{N^2}+\frac{4\gamma_2}{N}\frac{1}{f_2}+\frac{4\gamma_1^2}{N}\frac{1}{f_2},$$

$$M(S_2^3)=\frac{\gamma_4}{N^2}+\frac{12\gamma_2}{N}\frac{1}{f_2}+\frac{4\gamma_1^2}{N}\left(\frac{f_2-1}{f_2^2}-\frac{f_1}{f_2^2}\right)+\frac{8}{f_2^2}.$$

If the expansions are taken to order f^{-2} no other quantities than these will be required. Substituting into the formulae of the previous section and for brevity writing

$$\alpha_r=\frac{1}{f_1^r}-\frac{1}{f_2^r},\quad \beta_r=\frac{1}{f_1^r}+\frac{1}{f_2^r}$$

we have

$$\mathscr{E}(z)=-\tfrac{1}{2}(\alpha_1+\tfrac{1}{3}\alpha_2)+\frac{\alpha_1\gamma_2}{2\,N}+\frac{2}{3}\frac{\gamma_1^2}{N}\left(\alpha_1-\alpha_2+\frac{f_1}{f_2^2}\right),$$

$$\mathscr{E}(z^2)=\frac{1}{4}\left(2\beta_1+3\beta_2-\frac{2}{f_1f_2}\right)-\frac{\beta_1\gamma_2}{2\,N}-\frac{\gamma_1^2}{N}\left(\beta_1-\beta_2-\frac{f_1}{f_2^2}-\frac{1}{f_2}\right),$$

whence

$$\sigma_z^2=\frac{1}{2}\left(\frac{1}{f_1}+\frac{1}{f_2}\right)+\frac{1}{2}\left(\frac{1}{f_1^2}+\frac{1}{f_2^2}\right)-\frac{\gamma_2}{2(f_1+f_2+1)}\left(\frac{1}{f_1}+\frac{1}{f_2}\right)-\frac{\gamma_1^2}{f_1f_2^2}(f_2-f_1).$$

We note that when the populations generating the samples are normal, we obtain the correct first two terms of the expansion of σ_z^2 in terms of f_1 and f_2. It is clear, for the case considered, that the effect of the departure from normality of the parent population on the mean and the standard error of z will be slight and it is not surprising that Pearson (1931) was unable to draw any really firm conclusion from his sampling experiments.

14. These two illustrations are sufficient to show the great range of usefulness which the general expressions for the moments have. It has not been possible in the course of a single paper to do anything but the minimum of arithmetical investigation. It is hoped however to continue the arithmetical side in a further publication.

REFERENCES

DAVID, F. N. (1949). *Biometrika*, **36**, 394.
DAVID, F. N. & KENDALL, M. G. (1949). *Biometrika*, **36**, 431.
GAYEN, A. K. (1950). *Biometrika*, **37**, 236.
GEARY, R. C. (1947). *Biometrika*, **34**, 209.
JOHNSON, N. L. (1949). *Biometrika*, **36**, 149.
PATNAIK, P. B. (1949). *Biometrika*, **36**, 202.
PEARSON, E. S. (1931). *Biometrika*, **23**, 119.
QUENSEL, C. E. (1947). *Skand. AktuarTidskr.* **30**, 44.
TANG, P. C. (1938). *Statist. Res. Mem.* **2**, 126.

EFFICIENCY OF THE METHOD OF MOMENTS AND THE GRAM-CHARLIER TYPE A DISTRIBUTION

By L. R. SHENTON, *College of Technology, Manchester*

1. Introduction

We propose to consider the efficiency of the method of moments applied to fitting the Gram-Charlier series of Type A. In practice, since the sampling variances of moments higher than the fourth are large, only the first four moments of the observed frequency data are in general used, so that we shall confine our attention to the Type A series taken as far as the term involving the fourth moment.

It has been pointed out by Fisher (1921, pp. 355–6) that in the region of the normal point the Pearson system of curves closely resembles the system for which the method of moments is efficient; for 80% efficiency this region is restricted to $2{\cdot}62 < \beta_2 < 3{\cdot}42, \beta_1 < 0{\cdot}1$. The question arises as to how the departures from normality expressed by the Gram-Charlier type distribution affect the efficiency of moment-estimation and how this compares* with the situation for the Pearson system.

To evaluate the efficiency of moment estimation for the Gram-Charlier series we shall use a slight modification of the method given in a previous paper on maximum likelihood and moments (Shenton, 1950†). It consists in developing a determinantal expansion, which, since its terms are positive, provides a lower bound to the efficiency; in particular cases when there is one parameter for estimation it is possible to set lower and upper bounds to the efficiency.

2. The expansion for the information determinant

The expansion for the efficiency E_f given in M.L. expression (16) is capable of a simpler form in the case when the frequency function has a polynomial factor, as for example is the case with Type A. For the information determinant is of the form

$$\Delta = \left| \int_a^b \frac{A_j(x)\, A_k(x)\, w(x)\, dx}{B(x)} \right| \quad (j, k = 1, 2, \ldots, n), \tag{1}$$

and it may be shown under fairly general conditions that

$$\Delta = (-)^n \lim_{s \to \infty} \begin{vmatrix} [0] & [\alpha]_{ns} \\ [\alpha]'_{ns} & [\beta]_{ss} \end{vmatrix} \div |\, [\beta]_{ss} \,|, \tag{2}$$

where (i)

$$[\alpha]_{ns} = \begin{bmatrix} {}_1\alpha_0 & {}_1\alpha_1 & \cdots & {}_1\alpha_s \\ {}_2\alpha_0 & {}_2\alpha_1 & \cdots & {}_2\alpha_s \\ \vdots & & & \\ {}_n\alpha_0 & {}_n\alpha_1 & \cdots & {}_n\alpha_s \end{bmatrix}, \tag{3}$$

and $[\alpha]'_{ns}$ is its transpose.

$$\left.\begin{aligned} \text{(ii)}\ {}_j\alpha_k &= \int_a^b p_k(x)\, A_j(x)\, w(x)\, dx \quad (j = 1, 2, \ldots, n; k = 0, 1, \ldots, s). \\ \beta_{jk} = \beta_{kj} &= \int_a^b p_j(x)\, p_k(x)\, B(x)\, w(x)\, dx \quad (j, k = 0, 1, \ldots), \end{aligned}\right\} \tag{4}$$

* I am indebted to Prof. M. G. Kendall for this suggestion.

† To be referred to as M.L.

and $p_j(x), j = 0, 1, \ldots$, is a polynomial in x of precise degree j. The expansion (2), apart from a factor depending on the covariance term, is a compact form of (16) in M.L. and can be shown to converge in certain cases by an appeal to Parseval's theorem. In particular it converges to Δ when $a = -\infty$, $b = \infty$, $A_j(x)$ is a polynomial in x, $w(x) = e^{-\frac{1}{2}x^2}/\sqrt{(2\pi)}$ and $B(x)$ is a polynomial always positive for real x. Since (2) is equivalent to M.L. (16), the expansion will give an increasing sequence as an approximation to Δ. It is important to remember that the polynomial $B(x)$ must be positive for x in (a, b). We must therefore consider the conditions under which this is true for the case of Type A.

3. THE ADMISSIBLE PARAMETER VALUES FOR TYPE A

We consider the frequency function

$$P_x dx = \left\{1 + \frac{a_3}{3!}H_3(x) + \frac{a_4}{4!}H_4(x)\right\} g(x)\, dx, \tag{5}$$

where $x = (X - m)/\sigma, \quad g(x) = e^{-\frac{1}{2}x^2}/\sqrt{(2\pi)}$

and $H_r(x) = e^{-\frac{1}{2}D_x^2}x^r$ is the rth Hermite polynomial with the property

$$\left.\begin{aligned} \int_{-\infty}^{\infty} H_r(x)\, H_s(x)\, g(x)\, dx &= r! \quad (r = s), \\ &= 0 \quad (r \neq s). \end{aligned}\right\} \tag{6}$$

Using $\frac{d}{dx}H_r(x) = rH_{r-1}(x)$ along with $x^r = e^{\frac{1}{2}D_x^2}H_r(x)$ the moments of (5) may be seen to be

$$\left.\begin{aligned} \mu_{2r} &= \frac{(2r)!}{2^r r!}\left\{1 + r(r-1)\frac{a_4}{3!}\right\} \quad (r = 1, 2, \ldots), \\ \mu_{2r-1} &= \frac{(2r-1)!}{2^{r-2}(r-2)!}\frac{a_3}{3!} \quad (r = 2, 3, \ldots), \end{aligned}\right\} \tag{7}$$

so that for the measures of skewness and kurtosis we have

$$\beta_1 = \mu_3^2/\mu_2^3 = a_3^2, \quad \beta_2 = \mu_4/\mu_2^2 = a_4 + 3$$
$$\gamma_1 = a_3, \qquad \gamma_2 = a_4.$$

Now a term in the information determinant such as $E\left(\frac{\partial}{\partial m}\log P_x\right)^2$ involves the quartic $B(x) = 24 + 4a_3H_3(x) + a_4H_4(x)$ in the denominator so that to apply (2) we must ensure that this is positive for $-\infty < x < \infty$. Writing

$$H = -a_3^2 - a_4^2, \qquad J = -24a_4^2 - 2a_4^3 - 24a_3^2 - 6a_3^2a_4,$$
$$I' = 24a_4 + 6a_4^2 + 12a_3^2, \quad \nabla = I'^3 - 27J^2,$$

the quartic will have four imaginary roots if $\nabla > 0$ and either $H > 0$ or $2HI' - 3a_4J > 0$. (See, for example, Burnside & Panton, 1881, pp. 116 and 187.) Hence $B(x)$ is positive for real x provided

$$(8a_4 + 2a_4^2 + 4a_3^2)^3 > (24a_4^2 + 2a_4^3 + 24a_3^2 + 6a_3^2a_4)^2, \tag{8}$$

and

$$4a_4^3 - a_4^4 + 4a_4a_3^2 > 3a_3^2a_4^2 + 4a_3^4.$$

Table 1

a_4	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0	1·5	2·0	2·5	3·0	3·5
a_3^2	0·026	0·070	0·123	0·180	0·242	0·306	0·370	0·435	0·499	0·562	0·845	1·039	1·100	0·990	0·658

The second condition implies that $0 < a_4 < 4$. Table 1 shows the restriction imposed by (8) on a_3^2 for given a_4 in $(0, 4)$. For later use we compare the scope of Type A under these restrictions with the Pearson system of curves, using a (γ_1^2, γ_2) diagram. The region of validity of Type A is that enclosed by the curved boundary and the γ_2 axis in Fig. 1. Superimposed on this is the corresponding division for the Pearson system (Pearson, 1945, diagram XXXV, p. 66). It will be seen that the Pearson curves corresponding to the admissible Type A curves are Type IV, Type VII (including the heterotypic cases of these) and Types V and VI, putting them in approximate order of importance. Type IV and the symmetrical case of this, Type VII, occupy most of the Type A region. We notice that Types I and III do not share the region.

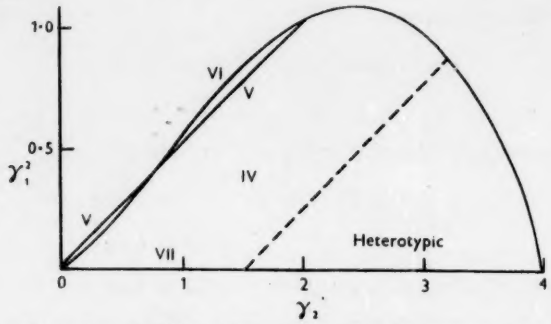


Fig. 1. Showing the region of validity of Type A (enclosed by curve and γ_2 axis) and the corresponding Pearson distribution.

4. Efficiency of the four parameter case of type A

The parameters m, σ, a_3, a_4 in (5) may be estimated by moments using

$$\overline{\theta}_1 = \Sigma' X/N, \quad \overline{\theta}_j = \Sigma'(X - \overline{\theta}_1)^j/N, \quad (j = 2, 3, 4).$$

The efficiency E_f is given by $\quad E_f^{-1} = |I| \times |[\mathrm{cov}\,(\overline{\theta}_j, \overline{\theta}_k)]|, \qquad (9)$

where the information matrix is $I = \left[E\left(\frac{N}{P_x^2}\frac{\partial P_x}{\partial \theta_j}\frac{\partial P_x}{\partial \theta_k}\right)\right]$.

Since to order N^{-1}

$$\mathrm{cov}\,(m_j, m_k) = \{\mu_{j+k} - \mu_j\mu_k + jk\mu_2\mu_{j-1}\mu_{k-1} - j\mu_{j-1}\mu_{k+1} - k\mu_{j+1}\mu_{k-1}\}/N,$$

using (7) it may be verified that

$$N^4\,|[\mathrm{cov}\,(\overline{\theta}_j, \overline{\theta}_k)]| = \sigma^{20}\begin{vmatrix} 2 - a_3^2 + a_4 & 6a_3 - a_3a_4 & -8 + 4a_4 \\ 6a_3 - a_3a_4 & 6 - a_3^2 + 9a_4 - a_4^2 & 12a_3 - a_3a_4 \\ -8 + 4a_4 & 12a_3 - a_3a_4 & 56 + 24a_4 - a_4^2 \end{vmatrix} = \sigma^{20} V \text{ say.} \quad (10)$$

For the information matrix I we take in (1)–(4)

$$A_j(x)\,g(x) = \frac{\partial P_x}{\partial \theta_j} \quad (j = 1, 2, 3, 4);$$

$$B(x)\,g(x) = P_x;$$

$$w(x) = g(x);$$

$$P_k(x) = H_k(x), \quad (k = 0, 1, \ldots);$$

$$a = -\infty, \quad b = \infty;$$

so that
$$ {}_j\alpha_k = \int_{-\infty}^{\infty} H_k(x) \frac{\partial P_x}{\partial \theta_j} dx, \quad (j = 1, 2, 3, 4;\ k = 0, 1, \ldots). \tag{11} $$

$$ \beta_{jk} = \beta_{kj} = \int_{-\infty}^{\infty} H_j(x)\, H_k(x)\, P_x dx, \quad (j, k = 0, 1, \ldots). \tag{12} $$

The integrals in (12) may be evaluated by using the recurrence relation

$$ H_j(x)\, H_k(x) = H_{j+k}(x) + jk H_{j+k-2}(x) + \frac{j(j-1)\, k(k-1)}{2!} H_{j+k-4}(x) + \frac{j(j-1)(j-2)\, k(k-1)(k-2)}{3!} H_{j+k-6}(x) + \ldots, $$

the last term on the right being a multiple of $H_1(x)$ or $H_0(x)$ according as $j+k$ is odd or even. Moreover,

$$ \frac{\partial P_x}{\partial \theta_1} = \left\{ H_1(x) + \frac{a_3}{3!} H_3(x) + \frac{a_4}{4!} H_5(x) \right\} \frac{g(x)}{\sigma}, $$

$$ \frac{\partial P_x}{\partial \theta_2} = \left\{ H_2(x) - \frac{H_4(x)}{2} + \frac{a_3}{3!} H_5(x) + \frac{a_4}{4!} H_6(x) \right\} \frac{g(x)}{2\sigma^2}, $$

$$ \frac{\partial P_x}{\partial \theta_3} = \frac{H_3(x)\, g(x)}{3!\ \sigma^3}, $$

$$ \frac{\partial P_x}{\partial \theta_4} = \frac{H_4(x)\, g(x)}{4!\ \sigma^4}. $$

Hence from (11) and (12) we find

$$ \left.\begin{aligned} &{}_1\alpha_1 = \frac{1}{\sigma}, \quad {}_1\alpha_4 = \frac{4a_3}{\sigma}, \quad {}_1\alpha_5 = \frac{5a_4}{\sigma}, \quad {}_1\alpha_j = 0, \quad (j \neq 1, 4, 5); \\ &{}_2\alpha_2 = \frac{1}{\sigma^2}, \quad {}_2\alpha_4 = -\frac{3!}{\sigma^2}, \quad {}_2\sigma_5 = \frac{10a_3}{\sigma^2}, \quad {}_2\alpha_6 = \frac{15a_4}{\sigma^2}, \quad {}_2\alpha_j = 0, \quad (j \neq 2, 4, 5, 6); \\ &{}_3\alpha_3 = \frac{1}{\sigma^3}, \quad {}_3\alpha_j = 0, \quad (j \neq 3); \\ &{}_4\alpha_4 = \frac{1}{\sigma^4}, \quad {}_4\alpha_j = 0, \quad (j \neq 4); \end{aligned} \right\} \tag{13} $$

and

$$ \left.\begin{aligned} \beta_{j,j} &= \left(1 + j(j-1)\frac{a_4}{4}\right) j! \equiv b_j, \\ \beta_{j,j+1} &= j(j+1)!\frac{a_3}{2} \equiv c_j, \\ \beta_{j,j+2} &= j(j+2)!\frac{a_4}{3!} \equiv d_j, \\ \beta_{j,j+3} &= (j+3)!\frac{a_3}{3!} \equiv f_j, \\ \beta_{j,j+4} &= (j+4)!\frac{a_4}{4!} \equiv g_j, \quad (j = 0, 1, \ldots); \\ \beta_{j,j+k} &= 0, \quad (k = 5, 6, \ldots). \end{aligned} \right\} \tag{14} $$

Substituting in (2) and eliminating as many of the border elements as possible we are led to

$$ \frac{\sigma^{20}}{N^4} |I| = \lim_{s \to \infty} \frac{K_{s-5}(B_5, C_5, d_5, f_5, g_5)}{K_s(b_0, c_0, d_0, f_0, g_0)}, \tag{15} $$

with
$$\left.\begin{aligned} B_5 &= b_5 - 20a_3f_2 - 10a_4g_1 + 25a_4^2b_1 + 100a_3^2b_2 + 100a_3a_4c_1 \\ &= 120 + 600a_4 - 200a_3^2 + 200a_3^2a_4 - 25a_4^2, \\ C_5 &= c_5 - 10a_3g_2 - 15a_4f_2 + 75a_4^2c_1 + 150a_3a_4b_2 \\ &= 1800a_3 + 225a_3a_4^2 - 300a_3a_4, \\ B_6 &= b_6 - 300a_4g_2 + 225a_4^2b_2 \\ &= 720 + 5400a_4 - 450a_4^2 + 225a_4^3, \\ B_s &= b_s, \quad s > 6, \quad C_s = c_s, \quad s > 5, \end{aligned}\right\} \tag{16}$$

where $K_s(p_k, q_l, r_m, \ldots)$ is a symmetric determinant of order s with elements

p_j, $(j = k, k+1, k+2, \ldots)$, in the diagonal through $(1,1)$,

q_j, $(j = l, l+1, l+2, \ldots)$, in the diagonal through $(2,1)$, and so on,

all unspecified elements being zero.* Since we may write V given by (10) as $K_5(b_0, c_0, d_0, f_0, g_0)$ we have for the joint efficiency of moment estimates for the four parameters of Type A in large sampling

$$E_f \leqslant \frac{K_s(b_0, c_0, d_0, f_0, g_0)}{K_5(b_0, c_0, d_0, f_0, g_0)\, K_{s-5}(B_5, C_5, d_5, f_5, g_5)}, \tag{17}$$

the ratio forming a decreasing sequence and the equality sign being applicable in the limit as $s \to \infty$. Thus for a particular value of s, (17) gives an upper bound for E_f. For computational purposes we may use $s = 6$ or $s = 7$ in (17), and the determinants $K_j(b_0, c_0, d_0, f_0, g_0)$, $j = 5, 6, 7$ may be simplified to become the first $j-2$ rows and columns of the array (18) below:

$$\begin{bmatrix} 2+a_4-a_3^2 & 6a_3-a_3a_4 & -8+4a_4 & -10a_3 & 30a_4 \\ 6a_3-a_3a_4 & 6+9a_4-a_3^2-a_4^2 & 12a_3-a_3a_4 & -30+15a_4+5a_3^2 & 120a_3 \\ -8+4a_4 & 12a_3-a_3a_4 & 56+24a_4-a_4^2 & 100a_3+5a_3a_4 & 360a_4 \\ -10a_3 & -30+15a_4+5a_3^2 & 100a_3+5a_3a_4 & 270+225a_4-25a_3^2 & 1200a_3 \\ 30a_4 & 120a_3 & 360a_4 & 1200a_3 & 720+5400a_4 \end{bmatrix}. \tag{18}$$

Table 2. *Values of E_f in the joint estimation of four parameters for Type A*

a_3^2 \ a_4	0·05	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0	1·5	2·0
0·0	0·99	0·97	0·92	0·88	0·84	0·81	0·78	0·75	0·73	0·71	0·69	0·63	0·59
0·1	—	—	—	0·60	0·70	0·73	0·72	0·71	0·69	0·68	0·66	0·60	0·57
0·2	—	—	—	—	—	0·57	0·64	0·66	0·65	0·65	0·63	0·58	0·54
0·3	—	—	—	—	—	—	0·48	0·58	0·61	0·61	0·60	0·56	0·52
0·4	—	—	—	—	—	—	—	—	0·54	0·57	0·57	0·53	0·50
0·5	—	—	—	—	—	—	—	—	—	—	0·54	0·51	0·48
0·6	—	—	—	—	—	—	—	—	—	—	—	0·49	0·46
0·7	—	—	—	—	—	—	—	—	—	—	—	0·47	0·44
0·8	—	—	—	—	—	—	—	—	—	—	—	0·45	0·42
0·9	—	—	—	—	—	—	—	—	—	—	—	—	0·40
1·0	—	—	—	—	—	—	—	—	—	—	—	—	0·38

In Table 2 we give values of (17) for $s = 6$ for admissible values of the parameters a_3^2 and a_4 (or γ_1^2 and γ_2) so that E_f is certainly less than the figure shown. From Table 2 it is seen that in the vicinity of the normal point E_f approaches unity; this is also suggested from (17), since

* $K_s(b_0, c_0, d_0, f_0, g_0)$ is a symmetric determinant with elements in nine diagonals only; it may therefore be regarded as a generalized continuant and the notation is suggested by Muir's (1882, pp. 149–60) for a continuant determinant.

from (13) c, d, f, g tend to zero with a_3 and a_4. Type A is therefore similar to the Pearson system in this respect. Further, it appears that $E_f < 80\,\%$ for $0{\cdot}6 \leqslant a_4 < 4$ and $a_3^2 \geqslant 0{\cdot}1$ and low efficiency is to be expected in the vicinity of the Type A boundary shown in Fig. 1. Although the efficiency seems to increase for increasing a_4 (with $a_3^2 = 0{\cdot}1$ say) this may be accounted for by the degree of approximation involved. To calculate further terms of (17) becomes laborious, but the following additional values assist in fixing the approximate position of the bound for the 80 % efficiency contour: $a_4 = 0{\cdot}1$, $a_3^2 = 0{\cdot}01$, $E_f < 0{\cdot}8950$ with $s = 6$, $E_f < 0{\cdot}8763$ with $s = 7$: $a_4 = 0{\cdot}2$, $a_3^2 = 0{\cdot}05$, $E_f < 0{\cdot}7029$ with $s = 6$. However in the case when $a_3^2 = 0$ the computation may be carried further, since in this case the determinants in (17) now factorize (using Laplacian expansion) and we find

$$E_f \leqslant \frac{\overline{K}_5(b_0, d_0, g_0)\,\overline{K}_s(b_1, d_1, g_1)}{\overline{K}_3(b_0, d_0, g_0)\,\overline{K}_{s-3}(B_6, d_6, g_6)\,\overline{K}_2(b_1, d_1, g_1)\,\overline{K}_{s-2}(B_5, d_5, g_5)}, \tag{19}$$

or

$$E_f \leqslant \frac{\overline{K}_{s+1}(b_0, d_0, g_0)\,\overline{K}_s(b_1, d_1, g_1)}{\overline{K}_3(b_0, d_0, g_0)\,\overline{K}_{s-2}(B_6, d_6, g_6)\,\overline{K}_2(b_1, d_1, g_1)\,\overline{K}_{s-2}(B_5, d_5, g_5)}, \tag{20}$$

where

$$B_s = b_s, \quad s > 6$$

$$B_5 = 120 + 600a_4 - 25a_4^2, \quad B_6 = 720 + 5400a_4 - 450a_4^2$$

and b, d, g are given in (14). The two expressions arise according as s is even or odd in (17). We use the notation $\overline{K}_s(p_k, q_l, \ldots)$ to indicate a symmetric determinant of order s with elements

p_j, $(j = k, k+2, k+4, \ldots)$, in the diagonal through $(1, 1)$,

q_j, $(j = l, l+2, l+4, \ldots)$, in the diagonal through $(2, 1)$, and so on,

all unspecified elements being zero. Thus, for example,

$$\overline{K}_3(b_0, d_0, g_0) = \begin{vmatrix} b_0 & d_0 & g_0 \\ d_0 & b_2 & d_2 \\ g_0 & d_2 & b_4 \end{vmatrix}, \qquad \overline{K}_2(B_5, d_5, g_5) = \begin{vmatrix} B_5 & d_5 \\ d_5 & b_7 \end{vmatrix}.$$

(19) and (20) give upper bounds for E_f and decreasing sequences and taking $s = 3, 4, 5$ we have the values in Table 3.

Table 3. *Successive values for E_f in the joint estimation of four parameters for Type A in the case $a_3 = 0$*

a_4	0·05	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0	1·5	2·0
1st approx.	0·99	0·97	0·92	0·88	0·84	0·81	0·78	0·75	0·73	0·71	0·69	0·63	0·59
2nd approx.	0·97	0·92	0·81	0·72	0·65	0·59	0·55	0·51	0·48	0·45	0·43	0·34	0·28
3rd approx.	0·97	0·92	0·81	0·72	0·65	0·59	0·54	0·50	0·46	0·43	0·41	0·31	0·25
4th approx.	0·97	0·92	0·81	0·71	0·63	0·57	0·52	0·47	0·43	0·40	0·37	0·26	0·19
5th approx.	0·97	0·91	0·80	0·71	0·63	0·57	0·52	0·47	0·43	0·40	0·37	0·26	0·19
6th approx.	0·97	0·91	0·80	0·71	0·63	0·57	0·52	0·47	0·43	0·40	0·37	0·25	0·17

The values in Table 3 are given correct to two figures; the first column of values correct to four figures reads: 0·9908, 0·9703, 0·9695, 0·9688, 0·9685 and 0·9680. The capricious behaviour of the successive approximations is noteworthy, so that although two successive terms may be nearly equal this provides little basis for supposing the remainder negligible. From Tables 2 and 3 it can now be seen that $E_f < 80\,\%$ for $0{\cdot}2 < a_4 < 4{\cdot}0$, $a_3^2 \geqslant 0{\cdot}1$, and since

$E_f < 0{\cdot}7029$ for $a_4 = 0{\cdot}2$, $a_3^2 = 0{\cdot}05$, it would appear that the 80 % efficiency contour is not in the region $a_3^2 \geqslant 0{\cdot}05$, $0{\cdot}2 \leqslant a_4 < 4{\cdot}0$. This may be compared with the case for Pearson's Type IV distribution, except that the general case is one of some complexity. Taking Type IV in the form

$$f(X)\,dX = y_0\{1+x^2\}^{-\frac{1}{2}(r+2)}\,e^{-\nu\tan^{-1}x}\,dx, \tag{21}$$

where $$x = \frac{X-m}{c}, \quad y_0 = \frac{1}{cF(r,\nu)}, \quad F(r,\nu) = e^{-\nu\pi}\int_0^{\pi}\sin^r\theta\, e^{\nu\theta}\,d\theta,$$

the moments following the well-known recurrence relation

$$\mu_s = -\frac{2\nu c(s-1)}{r(r-s+1)}\mu_{s-1} + \frac{(s-1)\,c^2(\nu^2+r^2)}{r^2(r-s+1)}\mu_{s-2}, \tag{22}$$

with $\mu_0 = 1$, $\mu_1 = 0$, $s \geqslant 2$; we have for the other moments up to μ_8:

$$\left.\begin{aligned}
\mu_2 &= c^2R^2/r^{(2)},\\
\mu_3 &= -4c^3\nu R^2/r^2r^{(3)},\\
\mu_4 &= 3c^4R^2\{(r-2)\,R^2+8\nu^2\}/r^3r^{(4)},\\
\mu_5 &= -8c^5\nu R^2\{(5r-12)\,R^2+24\nu^2\}/r^4r^{(5)},\\
\mu_6 &= 5c^6R^2\{3(r-2)\,(r-4)\,R^4+8(13r-36)\,R^2\nu^2+384\nu^4\}/r^5r^{(6)},\\
\mu_7 &= -12c^7\nu R^2\{(35r^2-238r+360)\,R^4+8(77r-240)\,R^2\nu^2+1920\nu^4\}/r^6r^{(7)},\\
\mu_8 &= 7c^8R^2\{15(r-2)\,(r-4)\,(r-6)\,R^6+8(170r^2-1259r+2160)\,R^4\nu^2\\
&\qquad +576(29r-100)\,R^2\nu^4+46080\nu^6\}/r^7r^{(8)};
\end{aligned}\right\} \tag{23}$$

where $$r^{(s)} \equiv r(r-1)\ldots(r-s+1) \quad \text{and} \quad R^2 = r^2+\nu^2.$$

The information matrix, given by Fisher (1921) is

$$I = \begin{bmatrix}
(r+1)(r+2)(r+4)\,A & -(r+1)(r+2)\,\nu A & (r+1)(r+2)\,B & -(r+1)\,\nu B\\
-(r+1)(r+2)\,\nu A & (r+1)(2r+8+\nu^2)\,A & -(r+1)\,\nu B & (r+2+\nu^2)\,B\\
(r+1)(r+2)\,B & -(r+1)\,\nu B & \dfrac{\partial^2}{\partial\nu^2}\log F & \dfrac{\partial^2}{\partial\nu\,\partial r}\log F\\
-(r+1)\,\nu B & (r+2+\nu^2)\,B & \dfrac{\partial^2}{\partial\nu\,\partial r}\log F & \dfrac{\partial^2}{\partial r^2}\log F
\end{bmatrix}, \tag{24}$$

where $$A^{-1} = c^2\{(r+4)^2+\nu^2\}, \quad B^{-1} = c\{(r+2)^2+\nu^2\}$$

It is evident from the form of the higher moments and the derivatives in I that E_f for joint estimation would be rather complicated. However, a comparison with Type A may be made in the special case $\nu = 0$. This situation may arise when we set out to fit a Type IV curve to a sample from a population in which $\nu = 0$, and where this fact is not known before the sample is drawn. The case should be distinguished from the one of simultaneous estimation of three parameters in a Type VII distribution which we consider in §5.

Using (23) we find

$$N^4\,|\,[\mathrm{cov}\,(m_r, m_s)]\,| = \frac{288c^{20}r^3(r-2)}{(r-1)^5\,(r-3)^5\,(r-5)^3\,(r-7)}, \tag{25}$$

$$\left.\frac{\partial(\mu_1,\mu_2,\mu_3,\mu_4)}{\partial(m,c,\nu,r)}\right|_{\nu=0} = \frac{48c^8}{r(r-1)^4\,(r-2)\,(r-3)^2}. \tag{26}$$

From (24) we have $$|\,I\,| = \frac{PQ}{4(r+2)^4\,(r+4)^2\,c^4}, \tag{27}$$

where $P = (r+2)^3 F(\frac{1}{2}r) - 2(r+1)(r+4)$,

$$Q = (r+1)^2(r+2)^2\{F[\tfrac{1}{2}(r-1)] - F(\tfrac{1}{2}r)\} - 2(r+1)(r+4),$$

$F(x) \equiv \dfrac{d^2}{dx^2}\log x! \equiv \psi'(x+1)$ in the notation of the trigamma function.*

Hence using (25)–(27) we have for the efficiency of the method of moments in the joint estimation of the four parameters of Type IV

$$E_f = \frac{32(r-3)(r-5)^3(r-7)(r+2)^4(r+4)^2}{r^5(r-1)^3(r-2)^3 PQ} \quad \text{when } \nu = 0. \tag{28}$$

Table 4. *Comparison of E_f for Type A and Type IV in the special case $\gamma_1 = 0$*

γ_2	0·05	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0	1·5
Type IV	0·99	0·98	0·92	0·84	0·74	0·63	0·53	0·43	0·33	0·25	0·18	0
Type A	0·97	0·91	0·80	0·71	0·63	0·57	0·52	0·47	0·43	0·40	0·37	0·25

For comparison with the corresponding case for Type A we must have $r = 3 + 6/a_4$ and the values of E_f are shown in Table 4. Since the entry for Type A is an upper bound we may conclude that higher efficiency occurs whenever $0 < \gamma_2 \leqslant 0{\cdot}6$ with moment estimation for Type IV than for Type A, and for $\gamma_2 > 0{\cdot}6$ there may be cases where the inefficiency for Type A is not as high as for Type IV. Type IV has $E_f < 80\,\%$ for $0{\cdot}35 \leqslant \gamma_2 \leqslant 1{\cdot}5$ while, for Type A, E_f is certainly below 80 % for $0{\cdot}2 < \gamma_2 < 4{\cdot}0$.

5. The symmetrical Type A and comparison with Pearson's Type VII

We now consider the case of Type A when $a_3 = 0$. Writing

$$P_x dx = \left(1 + \frac{a_4}{4!}H_4(x)\right) g(x)\,dx, \quad x = \frac{X-m}{\sigma}, \tag{29}$$

we have

$$\left.\begin{aligned} \frac{\partial P_x}{\partial m} &= \left(H_1(x) + \frac{a_4}{4!}H_5(x)\right)\frac{g(x)}{\sigma}, \\ \frac{\partial P_x}{\partial \sigma} &= \left(H_2(x) + \frac{a_4}{4!}[4H_4(x) + H_6(x)]\right)\frac{g(x)}{\sigma^2}, \\ \frac{\partial P_x}{\partial a_4} &= \frac{H_4(x)\,g(x)}{4!\,\sigma}. \end{aligned}\right\} \tag{30}$$

It is evident from the fact that the first of the equations in (30) is an odd function of x that for maximum likelihood estimators $\hat{m}$ is uncorrelated with $\hat{\sigma}$ and $\hat{a}_4$. The efficiency of the method of moments in this case then can be considered as follows:

Case A. Efficiency of the mean $\overline{m}$ (or m_1).

Case B. (i) Efficiency of $\overline{\sigma}$ given a_4, where $\overline{\sigma}^2 = m_2$. (ii) Efficiency of $\overline{a}_4$ given σ.

Case C. Efficiency of $\overline{\sigma}$ and $\overline{a}_4$, where $\overline{\sigma}^2 = m_2$ and $a_4 + 3 = \dfrac{m_4}{m_2^2}$.

* The F notation is due to Pairman, *Tracts for Computers*, No. 1; the psi function notation is used by H. T. Davis, *Tables of the Higher Mathematical Functions*, Vol. II, Principia Press, Inc.

Here we have used m_r for the rth moment of the sample of N about its mean m_1. Cases A and C are the most important in practice. Case B (i) may have interest for it corresponds to the efficiency of the estimation of $\bar{c}$ when r is known for the Type VII curve ($\nu = 0$ in (21)). Jeffreys (1939*b*, p. 708) has indicated that with data of observational errors, when the sample size N is less than 500, the estimate of the index $(r+2)/2$ is unreliable and a value 4 may be used in such cases. The parallel case with the symmetrical Type A would be to take $a_4 = 2$. Case B (ii) is not likely to occur in practice.

Case A. Efficiency of the mean

We have $$\frac{1}{N \operatorname{var} \hat{m}} = E\left(\frac{1}{P_x}\frac{\partial P_x}{\partial m}\right)^2 \quad \text{and} \quad N \operatorname{var} \overline{m} = \sigma^2,$$

so that in (2) we take $A_1(x)\,g(x) = \partial P_x/\partial m$ and following our previous method find

$$E_f(\overline{m}) \leqslant \frac{\overline{K}_s(b_1, d_1, g_1)}{\overline{K}_{s-1}(B_3, D_3, g_3)}, \quad s \geqslant 1, \tag{31}$$

with $$B_5 = 120 + 600a_4 - 25a_4^2, \quad B_s = b_s, \quad s \neq 5;$$
$$D_3 = 60a_4 - 5a_4^2, \qquad D_s = d_s, \quad s \neq 3;$$

using the notation of (14), (19) and (20). But we also have

$$\frac{1}{E_f(\overline{m})} = \int_{-\infty}^{\infty} \frac{\{H_1(x) + (a_4/4!)\,H_5(x)\}^2\, g(x)}{1 + (a_4/4!)\,H_4(x)}\,dx$$
$$= 1 + \frac{2a_4}{3} - \frac{a_4(4-a_4)}{6}\int_{-\infty}^{\infty} \frac{H_1^2(x)\,g(x)}{1 + (a_4/4!)\,H_4(x)}\,dx,$$

which leads to $$E_f(\overline{m}) \geqslant \left\{1 + \frac{2a_4}{3} - \frac{a_4(4-a_4)}{6}\frac{\overline{K}_{s-1}(b_3, d_3, g_3)}{\overline{K}_s(b_1, d_1, g_1)}\right\}^{-1}, \tag{32}$$

in the same notation as (31). From (31) and (32) upper and lower bounds may be found for $E_f(\overline{m})$. If we wish to make a comparison with Pearson's Type VII we find from (21), (23) and (24) with $\nu = 0$, $E_f(\overline{m}) = (r-1)(r+4)/\{(r+1)(r+2)\}$. With $s = 5$ in (31) and (32) we thus find the values in Table 5. Interpolating in the table it is seen that $E_f(\overline{m}) \geqslant 80\,\%$ for $0 \leqslant \gamma_2 \leqslant 1{\cdot}5$, and $E_f(\overline{m}) < 50\,\%$ for $3{\cdot}0 \leqslant \gamma_2 < 4{\cdot}0$. For Type VII $E_f(\overline{m}) > 80\,\%$ for $0 \leqslant \gamma_2 < 6{\cdot}0$ and is certainly higher than is the case for Type A. It may be noted that for Type A, since var (median) $= 1/(4NO_r^2)$, where O_r is the median ordinate, $\{\text{var (median)}/\text{var}(\overline{m})\} = 32\pi/(8 + a_4)^2$ which is less than unity for $a_4 > 2$; thus for $a_4 = 3$, $0{\cdot}55 < E_f$ (median) $< 0{\cdot}66$. Hence for $a_4 > 2$ the median is a slightly better estimator than $\overline{m}$.

Table 5. *$E_f(\overline{m})$ for Type A and Type VII*

γ_2	0·05	0·2	0·4	0·6	0·8	1·0	1·5	2·0	3·0
Type A (i)*	0·9996	0·9944	0·9799	0·9580	0·9299	0·8965	0·7951	0·6795	0·4527
(ii)	0·9996	0·9944	0·9801	0·9589	0·9324	0·9018	0·8141	0·7209	0·5493
Type VII	0·9996	0·9950	0·9842	0·9714	0·9583	0·9455	0·9167	0·8929	0·8571

* (i) and (ii) refer to lower and upper bounds respectively.

Case B (i). *Efficiency of $\bar{\sigma}$ given a_4*

From (30) and (2) we have $$\frac{\sigma^2}{N \operatorname{var} \hat{\bar{\sigma}}} \geqslant \frac{4\bar{K}_{s-2}(B_4, D_4, g_4)}{\bar{K}_s(b_0, d_0, g_0)},$$

where
$$B_4 = b_4 - 4a_4 d_2 + 4a_4^2 b_2 - g_2^2 = 24 + 72a_4 - 25a_4^2 + 4a_4^3,$$
$$D_4 = d_4 - 2a_4 g_2 - 15a_4 d_2 + 30a_4^2 b_2 = 480a_4 - 120a_4^2 + 30a_4^3,$$
$$B_s = b_s, \quad s > 6, \quad D_s = d_s, \quad s > 4;$$

and the other symbols are those given in (14) and (20), so that since $4N \operatorname{var} \bar{\sigma} = (a_4 + 2)\sigma^2$, we have
$$E_f(\bar{\sigma} \mid a_4) \leqslant \frac{\bar{K}_s(b_0, d_0, g_0)}{\bar{K}_2(b_0, d_0, g_0)\, \bar{K}_{s-2}(B_4, D_4, g_4)}. \tag{33}$$

A lower bound may also be found, for from

$$\int_{-\infty}^{\infty} \frac{\{H_2(x) + (a_4/3!)\, H_4(x) + (a_4/4!)\, H_6(x)\}^2}{1 + (a_4/4!)\, H_4(x)} g(x)\, dx$$
$$= 2 + a_4 - \frac{a_4(4 - a_4)}{6} \int_{-\infty}^{\infty} \frac{\{H_2(x) + 1\}^2}{1 + (a_4/4!)\, H_4(x)} g(x)\, dx$$

there follows
$$E_f(\bar{\sigma} \mid a_4) \geqslant \frac{4}{2 + a_4} \left\{ 2 + a_4 - \frac{a_4(4 - a_4)}{6} \frac{\bar{K}_{s-1}(B_2, D_2, g_2)}{\bar{K}_s(b_0, d_0, g_0)} \right\}^{-1} \tag{34}$$

where
$$B_2 = b_2 + 4b_0 = 6 + a_4, \quad B_s = b_s, \quad s > 2,$$
$$D_2 = d_2 - 2g_0 = 6a_4, \quad D_s = d_s, \quad s > 2,$$

and the other symbols are those used in (33).

Table 6. *$E_f(\bar{\sigma} \mid a_4)$ for Type A and the corresponding case for Type VII*

γ_2	0·05	0·2	0·4	0·6	0·8	1·0	1·5	2·0	3·0
Type A (i)*	0·9993	0·9920	0·9747	0·9500	0·9173	0·8759	0·7369	0·5684	0·2747
(ii)	0·9993	0·9920	0·9752	0·9525	0·9243	0·8908	0·7880	0·6704	0·4514
Type VII	0·9992	0·9893	0·9649	0·9341	0·9006	0·8667	0·7857	0·7143	0·6000

* (i) and (ii) refer to lower and upper bounds respectively.

For the corresponding problem for Type VII when r is given, it follows since $\mu_2 = c^2/(r-1)$ that $N \operatorname{var} \bar{c} = c^2 r/\{2(r-3)\}$, and from (24) $N \operatorname{var} \hat{c} = c^2(r+4)/\{2(r+1)\}$. Hence for the efficiency of $\bar{c}$ given r we have $E_f(\bar{c} \mid r) = (r-3)(r+4)/\{r(r+1)\}$. The comparison with Type A is given in Table 6, taking $s = 6$ in (33) and (34). It will be seen that $E_f(\bar{\sigma} \mid a_4) > 80\,\%$ for $0 \leqslant \gamma_2 \leqslant 1{\cdot}0$ and $< 60\,\%$ for $2{\cdot}0 < \gamma_2 < 4{\cdot}0$ approximately. For $0 < \gamma_2 < 1{\cdot}5$, E_f is higher for Type A than for Type VII, but the difference is not marked. It may be noted that in the range $2{\cdot}0 < \gamma_2 < 4{\cdot}0$ the efficiency is in the region of 70 % or below in both cases and is unsatisfactory; for Type VII, however, it has been shown by Sichel (1949) that the method of frequency-moments has a highly satisfactory efficiency.

Case B (ii). *Efficiency of $\bar{a}_4$ given σ*

For the maximum likelihood estimator $\hat{a}_4$ we have

$$\frac{1}{N \operatorname{var} \hat{a}_4} = \frac{1}{a_4^2}\left\{\int_{-\infty}^{\infty} \frac{g(x)\,dx}{1+(a_4/4!)\,H_4(x)} - 1\right\}, \tag{35}$$

$$\geqslant \frac{1}{a_4^2}\left\{\frac{\overline{K}_s(b_2, d_2, g_2)}{\overline{K}_{s+1}(b_0, d_0, g_0)} - 1\right\}, \tag{36}$$

and also

$$\left.\begin{aligned} \frac{1}{N \operatorname{var} \hat{a}_4} &= \frac{1}{6a_4(4-a_4)}\left\{6 - \int_{-\infty}^{\infty} \frac{\{H_2(x)-2\}^2}{1+(a_4/4!)\,H_4(x)} g(x)\,dx\right\}, \\ &\leqslant \frac{1}{3a_4(4-a_4)}\left\{3 - 2\frac{\overline{K}_s(B_2', D_2', g_2)}{\overline{K}_{s+1}(b_0, d_0, g_0)}\right\}, \end{aligned}\right\} \tag{37}$$

where

$$B_2' = b_2 + b_0 = 3 + a_4, \quad B_s' = b_s, \quad s > 2;$$
$$D_2' = d_2 + g_0 = 9a_4, \quad D_s' = d_s, \quad s > 2;$$

and b, d, g are otherwise given in (14).

For the estimation of a_4 by moments we may use

$$(a)\ \bar{a}_4 + 3 = m_4/\sigma^4, \quad (b)\ \bar{a}_4' = (m_4 - 6m_2\sigma^2 + 3\sigma^4)/\sigma^4, \quad (c)\ \bar{a}_4'' + 3 = m_4/m_2^2,$$

in which $m_r = \Sigma(X - \bar{X})^r/N$, $r = 2, 4$, and in large sampling we have approximately

$$N \operatorname{var} \bar{a}_4 = 96 + 204a_4 - a_4^2,$$
$$N \operatorname{var} \bar{a}_4' = 24 + 72a_4 - a_4^2,$$
$$N \operatorname{var} \bar{a}_4'' = 24 + 72a_4 - 25a_4^2 + 4a_4^3.$$

It is evident then that $\bar{a}_4$ is a poor estimator and indeed $E_f(\bar{a}_4 \mid \sigma) \to \frac{1}{4}$ for small a_4. The efficiency for the other two is given in Table 7, using $s = 4$ in (36) and (37).

Table 7. *$E_f(\bar{a}_4' \mid \sigma)$ and $E_f(\bar{a}_4'' \mid \sigma)$ for Type A*

a_4	0·05	0·2	0·4	0·6	0·8	1·0	1·5	2·0	3·0
$E_f(\bar{a}_4' \mid \sigma)$ (i)*	0·9753	0·8358	0·6722	0·5511	0·4602	0·3895	0·2660	0·1843	0·0771
(ii)	0·9759	0·8456	0·6980	0·5905	0·5101	0·4475	0·3373	0·2635	0·1660
$E_f(\bar{a}_4'' \mid \sigma)$ (i)	0·9755	0·8565	0·7213	0·6237	0·5508	0·4934	0·3867	0·3023	0·1448
(ii)	0·9761	0·8666	0·7490	0·6682	0·6105	0·5668	0·4904	0·4321	0·3118

* (i) and (ii) refer to lower and upper bounds respectively.

Thus in either case the efficiency is below 80 % for $a_4 > 0{\cdot}3$ approximately, and although $\bar{a}_4''$ is a better estimator than $\bar{a}_4'$ it cannot be regarded as satisfactory for a departure from normality exceeding a_4 or $\gamma_2 = 0{\cdot}4$. In other words, in large sampling from a Type A population with significant leptokurtosis, for which the parameters of scale and location are known (or the scale parameter alone is known) the method of moments should in general only be used as a first approximation.

Case C. The efficiency of the joint estimators $\bar{\sigma}$ and $\bar{a}_4$

Without going into details it is found using the previous methods that

$$E_f(\bar{\sigma}, \bar{a}_4) \leqslant \frac{\overline{K}_{s+3}(b_0, d_0, g_0)}{\overline{K}_3(b_0, d_0, g_0)\,\overline{K}_s(B_6, d_6, g_6)}, \tag{38}$$

with

$$B_s = b_s, \quad s > 6,$$

the other symbols being given in (14) and (20). For the corresponding problem with Type VII, using (23) and (24),

$$E_f(\bar{c},\bar{r}) = \frac{6(r+4)(r+2)^2(r+1)(r-5)^2(r-7)}{r^2(r-1)^3(r-2)(r-3)Q}, \tag{39}$$

where Q is given in (27), and $r = 3+6/a_4$. Taking $s = 3$ in (38) we find the values given in Table 8. For Type A then $E_f(\bar{\sigma},\bar{a}_4) < 80\,\%$ for $0{\cdot}4 < a_4 < 4{\cdot}0$, and less than $50\,\%$ for $1{\cdot}5 \leqslant a_4 < 4{\cdot}0$. On the other hand, the efficiency for Type VII is higher than for Type A for $0 < a_4 \leqslant 0{\cdot}5$ and perhaps a little outside this range, since the Type A entries are upper bounds. Thus for the range $0 < \gamma_2 \leqslant 0{\cdot}5$ Type VII is nearer Fisher's system of maximum moment efficiency (the exponential of a quartic) than is Type A.

Table 8. *$E_f(\bar{\sigma},\bar{a}_4)$ for Type A and $E_f(\bar{c},\bar{r})$ for Type VII*

a_4	0·05	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0	1·5	2·0	3·0
Type A	0·978	0·942	0·870	0·809	0·758	0·715	0·677	0·643	0·613	0·585	0·558	0·441	0·337	0·163
Type VII	0·996	0·986	0·947	0·888	0·815	0·731	0·642	0·550	0·460	0·372	0·291	0·000	—	—

It may be worth noting that an upper bound may also be found for the efficiency, in joint estimation, of either $\bar{\sigma}$ or $\bar{a}_4$. In the case of a_4 we have

$$N \operatorname{var} \hat{a}_4 = E\left(\frac{1}{P}\frac{\partial P}{\partial \sigma}\right)^2 \Big/ \Delta,$$

where

$$\Delta = \begin{vmatrix} E\left(\frac{1}{P}\frac{\partial P}{\partial \sigma}\right)^2 & E\left(\frac{1}{P^2}\frac{\partial P}{\partial \sigma}\frac{\partial P}{\partial a_4}\right) \\ E\left(\frac{1}{P^2}\frac{\partial P}{\partial a_4}\frac{\partial P}{\partial \sigma}\right) & E\left(\frac{1}{P}\frac{\partial P}{\partial a_4}\right)^2 \end{vmatrix},$$

so that we find

$$E\left(\frac{1}{P}\frac{\partial P}{\partial \sigma}\right)^2 = \lim_{s\to\infty} \frac{4K_{s-2}(B_4, D_4, g_4)}{\sigma^2 \bar{K}_s(b_0, d_0, g_0)}, \tag{40}$$

and

$$\Delta = \lim_{s\to\infty} \frac{4\bar{K}_{s-3}(B_6, d_6, g_6)}{\sigma^2 \bar{K}_s(b_0, d_0, g_0)} \tag{41}$$

in the notation of (14) and (33). Moreover, since $\bar{a}_4 + 3 = m_4/m_2^2$ so that

$$N \operatorname{var} \bar{a}_4 = 24 + 72a_4 - 25a_4^2 + 4a_4^3 = B_4$$

we find

$$E_f(\bar{a}_4) = \lim_{s\to\infty}\left\{\frac{\bar{K}_{s-2}(B_4, D_4, g_4)}{B_4 \bar{K}_s(b_0, d_0, g_0)}\right\} \lim_{s'\to\infty}\left\{\frac{\bar{K}_{s'}(b_0, d_0, g_0)}{\bar{K}_{s'-3}(B_6, d_6, g_6)}\right\} \tag{42}$$

$$= \lim_{s\to\infty} \frac{\bar{K}_s(B_4, D_4, g_4)}{\bar{K}_1(B_4)\,\bar{K}_{s-1}(B_6, d_6, g_6)}, \tag{43}$$

since neither of the limits in (42) is zero. Further (43) may be written

$$\begin{vmatrix} B_4 & D_4 & g_4 & \cdots \\ D_4 & B_6 & d_6 & \cdots \\ g_4 & d_6 & b_8 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix}_s \div \begin{vmatrix} 0 & B_4 & 0 & 0 & \cdots \\ 1 & B_4 & D_4 & g_4 & \cdots \\ 0 & D_4 & B_6 & d_6 & \cdots \\ 0 & g_4 & d_6 & b_8 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}_{s+1},$$

and so is a Schwein's determinantal ratio inverted; since the $\overline{K}_s(B_4, D_4, g_4)$ are essentially positive (43) therefore gives a decreasing sequence. As an illustration of this feature the values of $E_f(\overline{a}_4)$ for $a_4 = 1$ and $s = 1, 2, 3$ and 4 are 1·0000, 0·6560, 0·6165 and 0·6158. When $s = 4$ the values of $E_f(\overline{a}_4)$ are given in Table 9. The last row gives the corresponding value for Type VII found from (23) and (24) as

$$E_f(\overline{r}) = \frac{6(r+2)^2(r+1)^2(r-5)(r-7)}{r(r-1)^2(r-3)(r^2-r+18)Q}.$$

For Type A it is seen that $E_f(\overline{a}_4) < 80\,\%$ for $0{\cdot}4 < \gamma_2 < 4{\cdot}0$ and less than 50 % for $2{\cdot}0 < \gamma_2 < 4{\cdot}0$ approximately. $E_f(\overline{r})$ for Type VII is higher in $0 < \gamma_2 \leqslant 0{\cdot}4$.

Table 9. *Efficiency of $\overline{a}_4$ in joint estimation*

a_4	0·05	0·2	0·4	0·6	0·8	1·0	1·5	2·0	3·0
Type A	0·9782	0·8714	0·7665	0·6976	0·6503	0·6158	0·5556	0·5026	0·3625
Type VII	0·9964	0·9474	0·8181	0·6519	0·4777	0·3130	0·0000	—	—

6. The maximum likelihood equations

Considering the symmetrical Type A distribution for which

$$P_x dx = \left\{1 + \frac{a_4}{4!}H_4(x)\right\} g(x)\,dx, \quad x = \frac{X-m}{\sigma},$$

and writing for the likelihood function of a sample $(X_1, X_2, \ldots, X_N)$

$$L = \prod_{j=1}^{N} P_{xj}$$

we have, after simplification,

$$\frac{\partial \log L}{\partial m} = \Sigma\left\{H_1(x) - \frac{a_4 H_3(x)}{3!\,\Omega(x)}\right\}, \tag{44}$$

$$\frac{\partial \log L}{\partial \sigma} = \frac{1}{\sigma}\Sigma\left\{H_2(x) - 4 + \frac{8 - a_4 H_2(x)}{2\Omega(x)}\right\}, \tag{45}$$

$$\frac{\partial \log L}{\partial a_4} = \frac{1}{a_4}\Sigma\left\{1 - \frac{1}{\Omega(x)}\right\}, \tag{46}$$

in which $\Omega(x) = 1 + \frac{a_4}{4!}H_4(x)$ and Σ denotes summation over the sample. For maximum likelihood estimates we require solutions of these expressions equated to zero. Using $\overline{m}, \overline{\sigma}, \overline{a}_4$, the moment estimates as first approximations, we may find improved estimates $\hat{m}, \hat{\sigma}, \hat{a}_4$ from

$$\frac{\partial \log L}{\partial \overline{m}} + (\hat{m} - \overline{m})\frac{\partial^2 \log L}{\partial \overline{m}^2} = 0, \tag{47}$$

$$\frac{\partial \log L}{\partial \overline{\sigma}} + (\hat{\sigma} - \overline{\sigma})\frac{\partial^2 \log L}{\partial \overline{\sigma}^2} + (\hat{a}_4 - \overline{a}_4)\frac{\partial^2 \log L}{\partial \overline{\sigma}\,\partial \overline{a}_4} = 0, \tag{48}$$

$$\frac{\partial \log L}{\partial \overline{a}_4} + (\hat{\sigma} - \overline{\sigma})\frac{\partial^2 \log L}{\partial \overline{a}_4\,\partial \overline{\sigma}} + (\hat{a}_4 - \overline{a}_4)\frac{\partial^2 \log L}{\partial \overline{a}_4^2} = 0, \tag{49}$$

where

$$\frac{\partial^2 \log L}{\partial \overline{m}^2} = -\frac{1}{\overline{\sigma}} \Sigma \left\{ 1 + \frac{\overline{a}_4(4+H_2(x))}{3!\,\Omega(x)} - \frac{\overline{a}_4(4-\overline{a}_4)\,(1+H_2(x))}{3!\,\Omega^2(x)} \right\},$$

$$\frac{\partial^2 \log L}{\partial \overline{\sigma}^2} = \frac{1}{\overline{\sigma}^2} \Sigma \left\{ 2 - 3H_2(x) + \frac{(24-10\overline{a}_4-\overline{a}_4 H_2(x))}{2\Omega(x)} - \frac{(4-\overline{a}_4)\,(8-\overline{a}_4-2\overline{a}_4 H_2(x))}{2\Omega^2(x)} \right\},$$

$$\frac{\partial^2 \log L}{\partial \overline{a}_4 \partial \overline{\sigma}} = -\frac{4}{\overline{a}_4 \overline{\sigma}} \Sigma \left\{ \frac{1}{\Omega(x)} - \frac{(8-\overline{a}_4 H_2(x))}{\Omega^2(x)} \right\},$$

$$\frac{\partial^2 \log L}{\partial \overline{a}_4^2} = -\frac{1}{\overline{a}_4^2} \Sigma \left\{ 1 - \frac{1}{\Omega(x)} \right\}^2,$$

and

$$\Omega(x) \equiv \Omega(x, \overline{m}, \overline{\sigma}),$$

and in which cross-product terms involving m such as $\partial^2 \log L/\partial m\,\partial\sigma$ are ignored since their expectations are zero. In the case when $a_4 \leqslant 1{\cdot}5$ approximately we see from Table 5 that $\overline{m}$ is an estimate of m with 80 % or more efficiency in large sampling. Hence when $a_4 \leqslant 1{\cdot}5$ improved estimates of σ and a_4 may be found from $\overline{\sigma}$ and $\overline{a}_4$ using (48) and (49) only. The case when $a_4 > 1{\cdot}5$ is likely to be more involved. For this we could use (47) to improve $\overline{m}$ (or the median estimate of m) and (48)–(49) to improve $\overline{\sigma}$ and $\overline{a}_4$. The process could be repeated if necessary.

The parallel problem with Pearson's Type VII has been discussed by Jeffreys (1939*b*; 1948, pp. 184–7), who uses a different form from (21) with $\nu = 0$, involving orthogonal parameters θ_j, $j = 1, 2, 3$, in the sense that the expectation of such terms as $\partial^2 \log L/\partial\theta_1\,\partial\theta_2$ is zero. A different approach to the problem has been described by Sichel (1947, 1949), whose method of frequency-moments is fairly efficient in the range where the method of moments is very inefficient. It would seem that if we are to find efficient estimates of the parameters involved, then the use of Type VII may be less laborious than Type A. It is clear that the solution of the likelihood equations for an asymmetrical Type A distribution would be no light undertaking.

7. SUMMARY

The results found in this paper may be conveniently summarized as follows:

(*a*) For the Type A distribution of four parameters (given by (5)) the joint efficiency of the method of moments is less than 80 % for $0{\cdot}2 < \gamma_2 < 4{\cdot}0$, $\gamma_1^2 \geqslant 0{\cdot}05$; and falls rapidly with increasing skewness and leptokurtosis, and also in the vicinity of critical values of the parameters a_3, a_4 which arise from the condition that the frequency must be positive. For the corresponding problem for Pearson's Type IV we have shown that the efficiency E_f is below 80 % for $0{\cdot}35 \leqslant \gamma_2 < 1{\cdot}5$, $\gamma_1^2 = 0$ and above 80 % for $0 \leqslant \gamma_2 < 0{\cdot}35$, $\gamma_1^2 = 0$.

(*b*) For the symmetrical Type A

$$P_x = \left\{ 1 + \frac{a_4}{4!} H_4(x) \right\} e^{-\frac{1}{2}x^2}/\sqrt{(2\pi)}, \quad x = \frac{X-m}{\sigma},$$

and Pearson's Type VII

$$f(X) = \frac{(\frac{1}{2}r)!}{(\frac{1}{2}r-1)!} \frac{c^{r+1}}{\{c^2+(X-m)^2\}^{\frac{1}{2}(r+2)}}$$

we have:

Parameters estimated	$\overline{m}$	$\overline{\sigma}$ given a_4 $\overline{c}$ given r	$\overline{a}_4$ given σ	$\overline{\sigma}$ and $\overline{a}_4$ $\overline{c}$ and $\overline{r}$
$E_f < 80\%$ Type A	$1{\cdot}5 < \gamma_2 < 4{\cdot}0$	$1{\cdot}35 < \gamma_2 < 4{\cdot}0$	$0{\cdot}3 < \gamma_2 < 4{\cdot}0$	$0{\cdot}3 < \gamma_2 < 4{\cdot}0$
Type VII	$\gamma_2 > 6{\cdot}0$	$\gamma_2 > 1{\cdot}4$	—	$0{\cdot}42 \leqslant \gamma_2 < 1{\cdot}5$
$E_f > 80\%$ Type A	$0 \leqslant \gamma_2 \leqslant 1{\cdot}5$	$0 \leqslant \gamma_2 \leqslant 1{\cdot}35$	$0 \leqslant \gamma_2 < 0{\cdot}3$	—
Type VII	$0 \leqslant \gamma_2 \leqslant 6{\cdot}0$	$0 \leqslant \gamma_2 \leqslant 1{\cdot}4$	—	$0 \leqslant \gamma_2 < 0{\cdot}42$

(*c*) In the vicinity of the normal point ($0 < \gamma_2 < 0{\cdot}5$ approx.) the efficiency of the method of moments in joint estimation is higher for Type VII than for the symmetrical Type A; this remark also applies to Type IV and Type A with four parameters when $\gamma_1^2 = 0$. In both cases it appears that near the normal point efficiency is high. (*Note.* It has been shown by Fisher (1921) that the system of curves for which the method of moments gives maximum likelihood estimates is of the form $y = e^{A(x)}$, where $A(x) = bx^2 + cx^3 + dx^4$ and if c/b and d/b are small, satisfies a differential equation similar to Pearson's. From this point of view it might be anticipated that near the normal point the method of moments would have high efficiency for the Pearson system. But, as has been noted by Jeffreys (1939*a*), the differential equation of Fisher's curves cannot approximate to a Type IV unless d is positive, in which case y diverges ultimately. Furthermore, it appears that the Type A curves are not mathematical approximations near the normal point to the Fisher system and yet we have indicated that these have high efficiency. It may be that Type A and Type IV have approximately the same moments as the Fisher system near the normal point, after identification of the first four moments, which would explain the high efficiency.)

8. Conclusion

Except in the cases of slight departure from normality such as occur when $0 < \gamma_2 < 0{\cdot}3$ and $0 \leqslant \gamma_1^2 \leqslant 0{\cdot}05$, it appears that the method of moments is inefficient when applied to Type A in the form considered here. A departure from normality of this size would require a large sample to detect with any degree of certainty. Thus in the case of the symmetrical Type A with σ known, using (37) we have $N \operatorname{var} \hat{a}_4 > 3(4 - a_4)(2 + a_4)$, so that if a_4 is to exceed $3\sqrt{(\operatorname{var} \hat{a}_4)}$ then $N > 5700$ approximately when $a_4 = 0{\cdot}2$. In general, then, maximum likelihood estimates should be found whenever there is reason to believe the observed distribution arises from a Type A population.

It is possible that these remarks also apply to Edgeworth's form of Type A,

$$P_x = \left\{1 + \frac{a_3}{3!}H_3(x) + \frac{a_4}{4!}H_4(x) + \frac{10a_3^2}{6!}H_6(x)\right\} g(x).$$

It would not, however, be an easy matter to find the admissible parameter values for P_x to be positive in this case. But the method described here would lead, in the general case, to an upper bound for the efficiency of estimation by moments. This would not provide grounds for deciding whether Edgeworth's form was better from the point of view of efficiency, for we should be comparing upper bounds.

The exact bearing of our results on the practical problem of estimating parameters from a sample from a leptokurtic population is not easy to assess. For the measure of efficiency

calculated is based on a large sample limit and may be lower for a finite sample. Again for a Pearson distribution in the so-called heterotypic region the efficiency of estimation by moments may be zero. Thus in sampling from a Type VII population $y = y_0(c^2+x^2)^{-\frac{1}{2}(r+2)}$, we have for large samples $\text{var}\, m_2 = 2r/\{N(r-3)(r-1)^2\}$ which is infinite for $r = 3$, so that the efficiency of an estimate depending on the second moment would be zero. But in finite sampling in this case the variance of the second moment would not be infinite so that the efficiency would not be as low as the large sample figure suggests. A similar remark applies to estimation by moments in the case of Type A when the parameters lie in the vicinity of the critical region (a border strip on the curvilinear boundary of Fig. 1). For in this region we have shown that the efficiency is low or zero, but in practice it may be that the method of moments is not so useless.

The comparison between the Pearson Type VII and symmetrical Type A distributions having the same first four moments suggests that the method of moments is more efficient with the former, at least in the vicinity of the normal point. Further, in a situation in which we consider that an observed distribution can be described either by a Pearson or Type A curve, the shape of the distribution may be important. Now the Pearson system is unimodal but this is not the case in general with Type A. The form of Type A considered here may have three modes, and from a practical point of view statistical data showing such a characteristic might be regarded as heterogeneous. Thus if we are thinking in terms of unimodal distributions and do not allow negative frequencies the admissible parameter values for Type A will not be as extensive as is implied in Fig. 1.

The method we have described in this paper could be applied to other forms of Type A, such as those based on the logarithmically transformed normal curve or Pearson's Type III, provided the questions of convergence involved in (2) were settled. Discrete distributions such as Charlier's Type B could also be treated.

REFERENCES

BURNSIDE, W. S. & PANTON, A. W. (1881). *Theory of Equations*. Longmans.

FISHER, R. A. (1921). On the mathematical foundations of theoretical statistics. *Philos. Trans.* A, **222**, 309.

JEFFREYS, H. (1939*a*). The law of error and the combination of observations. *Philos. Trans.* A, **237**, 231.

JEFFREYS, H. (1939*b*). The law of error in the Greenwich variation of latitude observations. *Mon. Not. R. Astr. Soc.* **99**, 703.

JEFFREYS, H. (1948). *Theory of Probability*. Oxford: Clarendon Press.

MUIR, T. (1882). *A Treatise on the Theory of Determinants*. Macmillan and Co.

PEARSON, K. (1945). *Tables for Statisticians and Biometricians*, Part 1. Biometrika Office.

SHENTON, L. R. (1950). Maximum likelihood and the efficiency of the method of moments. *Biometrika*, **37**, 111.

SICHEL, H. S. (1947). Fitting growth and frequency curves by the method of frequency-moments. *J.R. Statist. Soc.* **110**, 337.

SICHEL, H. S. (1949). The method of frequency-moments and its application to Type VII populations. *Biometrika*, **36**, 404.

ON DISTRIBUTIONS FOR WHICH THE HARTLEY-KHAMIS SOLUTION OF THE MOMENT-PROBLEM IS EXACT

By H. P. MULHOLLAND

1. Introduction

Suppose that the moments of a certain distribution are known exactly up to order n (inclusive). Hartley & Khamis (1947) have described a procedure yielding estimated values of the cumulative distribution function (c.d.f.) at equally spaced points belonging to a chosen set or 'grid'; the resulting table for the c.d.f. has a rather wide tabular interval, h say, and subtabulation can be achieved by shifting the grid (Hartley & Khamis, 1947, §7, especially p. 348). The present paper is intended to give a clue to the type of case for which this procedure is suitable by determining the class of distributions that it reproduces exactly. (Owing to the basis of the procedure (Hartley & Khamis, 1947, §2) this happens when Sheppard's corrections for grouping are exact.) It will appear in §§3 and 4 (I), that the distributions in question are those generated by the procedure itself in the cases where its estimate for the c.d.f. is an increasing function, and that in any case its result is representable by a finite expansion fitted by moments and resembling the Gram-Charlier Type A expansion (carried as far as the given moments permit). Moreover, the leading term in the former expansion represents the n-fold convolution of a rectangular distribution with itself. Since this rapidly approaches normality as n increases, the resemblance to the Type A expansion is numerical as well as formal; it will be discussed further and illustrated numerically in §8. Altogether, *despite its apparent dissimilarity, the Hartley-Khamis procedure is akin to the various classical expansion methods, especially the Type A expansion.* We may thus expect the procedure to work best when the leading term in its expansion yields already a good first approximation. It is to be noted that the distribution given by this term, besides being nearly normal, is symmetrical, slightly platykurtic, and of finite range.

In §4 (II), certain classes of distributions are shown to have exact Hartley-Khamis estimates for *one* position of the grid. In §7 (*a*) alternative methods of subtabulation are compared, especially as regards continuity; in §7 (*b*) I give *rules for choosing the interval h*; and in §7 (*c*)–(*e*), it is noted that some of the formulae leading to (I) can serve as the starting-points of investigations into the magnitudes of the errors in the Hartley-Khamis estimates.

2. Notation and preliminaries

After a suitable change of scale and origin (cf. Hartley & Khamis, 1947, §3 (15), p. 343) we shall have h replaced by 1 and the centres of the grid-intervals or 'groups' at the points

$$x = \xi_i = i - \tfrac{1}{2}n \quad (i = \ldots, -1, 0, 1, \ldots, n, n+1, \ldots). \tag{2·1}$$

Supposing that the c.d.f. is now $P(x)$, we introduce its central difference

$$\overline{p}(x) = \delta P(x) = P(x+\tfrac{1}{2}) - P(x-\tfrac{1}{2}). \tag{2·2}$$

Then for the 'grouped frequencies' $\overline{p}(\xi_i)$ the Hartley-Khamis procedure provides estimates $\overline{p}^*(\xi_i)$ determined by the conditions

$$\sum_{i=0}^{n} \xi_i^r \overline{p}^*(\xi_i) = \int_{-\infty}^{\infty} x^r \overline{p}(x)\, dx = \overline{\mu}'_r \quad (r = 0, 1, \ldots, n), \tag{2·3}$$

$$\overline{p}^*(\xi_i) = 0 \quad (i = \ldots, -2, -1, \text{ or } i = n+1, n+2, \ldots). \tag{2·4}$$

In fact, the moments $\overline{\mu}'_r$ of $\overline{p}(x)$ about 0, when combined with Sheppard's corrections (for groups of unit width), yield the moments μ'_r about 0 belonging to the c.d.f. $P(x)$, so that equations (2·3) are equivalent to Hartley & Khamis's equations (13) (1947, p. 342; see also Kendall, (3·40), p. 70, and (3·43), (3·44), p. 71). The simultaneous equations (2·3) may be solved numerically for the numbers $\overline{p}^*(\xi_i)\,(i = 0, 1, \ldots, n)$, as is done by Hartley and Khamis. Alternatively, observing that, subject to (2·4), the equations (2·3) hold if and only if

$$\sum_{j=-\infty}^{\infty} \pi_n(\xi_j)\,\overline{p}^*(\xi_j) = \int_{-\infty}^{\infty} \pi_n(x)\,\overline{p}(x)\,dx \tag{2·5}$$

for an arbitrary polynomial $\pi_n(x)$ of degree not exceeding n, we may substitute for $\pi_n(x)$ a fundamental Lagrange interpolation polynomial $l_\xi(x)$ given by

$$l_\xi(x) = \phi_{n+1}(x)\,(x-\xi)^{-1}/\phi'_{n+1}(\xi) \quad (\xi = -\tfrac{1}{2}n, -\tfrac{1}{2}n+1, \ldots, \tfrac{1}{2}n), \tag{2·6}$$

$$\phi_{n+1}(x) = x^{(n+1)} = (x+\tfrac{1}{2}n)^{(n+1)} = (x+\tfrac{1}{2}n)\,(x+\tfrac{1}{2}n-1)\ldots(x-\tfrac{1}{2}n). \tag{2·7}$$

Since $l_\xi(\xi) = 1$ and $l_\xi(\xi_j) = 0\ (\xi_j \neq \xi)$, we get from (2·5)

$$\overline{p}^*(\xi) = \int_{-\infty}^{\infty} l_\xi(x)\,\overline{p}(x)\,dx \quad (\xi = \xi_i = i - \tfrac{1}{2}n, i = 0, 1, \ldots, n); \tag{2·8}$$

and, conversely, (2·8) and (2·4) imply (2·5), since any $\pi_n(x)$ is a linear combination of the polynomials (2·6). Further, the equality in (2·8) will hold, by (2·4), for the other values of ξ also if we write

$$l_\xi(x) \equiv 0 \quad (\xi = \xi_i = i - \tfrac{1}{2}n; i = \ldots, -2, -1, n+1, n+2, \ldots). \tag{2·9}$$

Similarly, if we take the grid symmetrical about η instead of about 0 we shall get

$$\overline{p}^*(\xi+\eta) = \int_{-\infty}^{\infty} l_\xi(x-\eta)\,\overline{p}(x)\,dx \quad (\xi = i - \tfrac{1}{2}n; i = \ldots, -1, 0, 1, 2, \ldots). \tag{2·10}$$

If η is allowed to vary continuously in the interval $-\frac{1}{2} \leqslant \eta < \frac{1}{2}$ *the function* $\overline{p}^*(x)$ *is now defined for all values of* x *by* (2·10),† and we may obtain a corresponding estimate $P^*(x)$ for $P(x)$ by summation:

$$P^*(x) = \sum_{j=-\infty}^{0} \overline{p}^*(x - \tfrac{1}{2} + j). \tag{2·11}$$

This series is only nominally infinite, in virtue of (2·9). Clearly $\overline{p}^*(x)$ is obtained from $\overline{p}(x)$ by a linear integral transformation

$$\overline{p}^*(x) = \int_{-\infty}^{\infty} K_n(x, y)\,\overline{p}(y)\,dy, \tag{2·12}$$

whose kernel $K_n(x, y)$ is given, according to (2·10), by

$$K_n(x, y) = l_\xi(\xi + y - x) \quad (-\tfrac{1}{2} \leqslant x - \xi < \tfrac{1}{2}, \xi = i - \tfrac{1}{2}n; i = \ldots, -1, 0, 1, 2, \ldots). \tag{2·13}$$

Evidently, for fixed x, $K_n(x, y)$ is a polynomial of degree n in y (or is zero). Thus we shall have, with ξ and x as in (2·13),

$$K_n(x, y) = \sum_{r=0}^{n} l_\xi^{(r)}(\xi - x)\,y^r/r!, \tag{2·14}$$

$$\overline{p}^*(x) = \sum_{r=0}^{n} l_\xi^{(r)}(\xi - x)\,\overline{\mu}'_r/r!. \tag{2·15}$$

When $x = \xi$ equation (2·15) corresponds to Hartley & Khamis's equation (14) (1947, p. 343), and, indeed, yields explicit formulae for the coefficients in their equation.

† For other ways of obtaining estimates for $\overline{p}(x)$ at values of x intermediate between those in (2·1) see §7 (*a*).

3. Preservation of moments and expansion properties

Let $g(x)$ be any function of bounded variation on $(-\infty, \infty)$ such that the absolute moments of $|g(x)|$ exist up to the nth order (inclusive), and let us transform it as in (2·12) so as to get

$$g^*(x) = \int_{-\infty}^{\infty} K_n(x, y)\, g(y)\, dy. \tag{3·1}$$

We note first that *this transformation preserves moments of orders not exceeding n.* In fact, from the construction of the transformation (cf. especially (2·3) and (2·10) above) we have

$$\mu_r'[g] = \int_{-\infty}^{\infty} x^r g(x)\, dx = \sum_{i=-\infty}^{\infty} (\xi_i + \eta)^r g^*(\xi_i + \eta) \quad (r = 0, 1, \ldots, n), \tag{3·2}$$

and on averaging both sides over the interval $-\frac{1}{2} \leqslant \eta < \frac{1}{2}$ we get

$$\mu_r'[g] = \sum_{-\infty}^{\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} (\xi_i + \eta)^r g^*(\xi_i + \eta)\, d\eta = \int_{-\infty}^{\infty} x^r g^*(x)\, dx = \mu_r'[g^*],$$

for $r = 0, 1, \ldots, n$, as required. (As in (2·11) the series are only nominally infinite.)

Secondly, since the transform of a function according to (3·1) depends only on its moments (of orders not exceeding n), it is clear from what has just been proved that *the transformation* (3·1) *is 'consistent', i.e. that* $g^{**}(x) = g^*(x)$. Let us note that this property is equivalent to the following relation for the 'iterated' kernel:

$$K_n^{(2)}(x, y) = \int_{-\infty}^{\infty} K_n(x, u)\, K_n(u, y)\, du = K_n(x, y). \tag{3·3}$$

(The range of integration is only nominally infinite on account of (2·9) and (2·13).)

Thirdly, *the kernel* $K_n(x, y)$ *possesses expansions in the form*

$$K_n(x, y) = \sum_{r=0}^{n} \alpha_r(x)\, \beta_r(y), \tag{3·4}$$

where the functions $\alpha_r(x)$ $(r = 0, 1, \ldots, n)$ *are linearly independent and* $\beta_r(y)$ *is a polynomial of degree exactly r: the kernel* $K_n(x, y)$ *is thus 'degenerate'.* In fact, the expansion (2·14) is clearly of the type (3·4), and other such expansions are obtainable (e.g. that in (5·1) below). Substituting the expansion (3·4) for K_n in (3·3) we find

$$\sum_{r=0}^{n} \sum_{s=0}^{n} \alpha_s(x)\, \beta_r(y) \int_{-\infty}^{\infty} \alpha_r(u)\, \beta_s(u)\, du = \sum_{r=0}^{n} \alpha_r(x)\, \beta_r(y).$$

Since the $(n+1)^2$ functions $\alpha_s(x)\, \beta_r(y)$ of x and y are clearly linearly independent we must have

$$\int_{-\infty}^{\infty} \alpha_r(u)\, \beta_s(u)\, du = \begin{cases} 0 & (r \neq s) \\ 1 & (r = s) \end{cases} \quad (r, s = 0, 1, \ldots, n), \tag{3·5}$$

i.e. the two sets of functions $\{\alpha_r(u)\}, \{\beta_s(u)\}$ *form a normalized biorthogonal system.*

Using (3·4) we get an expansion for $g^*(x)$ as follows:

$$g^*(x) = \int_{-\infty}^{\infty} \sum_{r=0}^{n} \alpha_r(x)\, \beta_r(y)\, g(y)\, dy = \sum_{r=0}^{n} g_r \alpha_r(x), \tag{3·6}$$

$$g_r = \int_{-\infty}^{\infty} \beta_r(y)\, g(y)\, dy. \tag{3·7}$$

Further, from (3·5) and (3·6) it follows that $g(x)$ *satisfies the integral equation* $g(x) = g^*(x)$, *i.e.*

$$g(x) = \int_{-\infty}^{\infty} K_n(x, y)\, g(y)\, dy, \tag{3·8}$$

if and only if it has the form

$$g(x) = \sum_{r=0}^{n} g_r \alpha_r(x) \tag{3·9}$$

with constant coefficients g_r.

4. MAIN RESULTS

From among the possible expansions of $K_n(x, y)$ in the form (3·4) we now select one in which *the leading term is continuous together with as many of its derivatives as possible.* The results that emerge are given in (I) and (II) below; in them the functions $P(x)$ (hitherto confined to c.d.f.'s) are subject only to the restrictions

$$\int_{-\infty}^{\infty} |x|^r |dP(x)| < \infty \quad (r = 0, 1, \ldots, n), \tag{4·1}$$

which when $P(x)$ is a c.d.f. assert the finiteness of its absolute moments of orders 0 to n.

(I) *With the notations of* §2 *and subject to* (4·1) *we have*

$$P^*(x) = P(x) \tag{4·2}$$

for all values of x *if and only if* $P(x)$ *is of the form*

$$P(x) = \sum_{r=0}^{n} a_r V_n^{(r)}(x) \tag{4·3}$$

with constant coefficients a_r, *where* $V_n(x)$ *is the c.d.f. of the sum of* n *independent variates each of which has the rectangular distribution on* $(-\frac{1}{2}, \frac{1}{2})$. *Further, in all cases we have*

$$P^*(x) = \sum_{r=0}^{n} a_r V_n^{(r)}(x) = \sum_{r=0}^{n} V_n^{(r)}(x) \int_{-\infty}^{\infty} \overline{w}_r(t)\, dP(t), \tag{4·4}$$

$$\overline{w}_r(t) = \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} w_r(y)\, dy, \quad w_r(y) = (-)^r \left(\frac{d}{dy}\right)^{n-r} \frac{y^{(n)}}{n!} \quad (r = 0, 1, \ldots, n), \tag{4·5}$$

In particular, $V_n(x)$ *is the only c.d.f. with a continuous* $(n-1)$*th derivative for which* (4·2) *holds identically.*

(II) *The equality* (4·2) *holds when* $x + \frac{1}{2}$ *belongs to the discrete set of points* (2·1) *if*

$$P(x) \equiv \sum_{r=0}^{n} b_r W_{n+1}^{(r)}(x), \tag{4·6}$$

where the b_r*'s are arbitrary constants and* $W_{n+1}(x)$ *is the c.d.f. of the sum of two independent variates, one with c.d.f.* $V_n(x)$ *and the other arbitrarily distributed on the open interval* $(-\frac{1}{2}, \frac{1}{2})$.

In particular, (4·2) *holds whenever* $x + \frac{1}{2}$ *is one of the points* (2·1) *if*

$$P(x) \equiv \sum_{r=0}^{n} c_r V_{n+1}^{(r)}(x), \tag{4·7}$$

where the c_r*'s are arbitrary constants; and, further, in all cases, if we write*

$$P_*(x) \equiv \sum_{r=0}^{n} c_r V_{n+1}^{(r)}(x) \equiv \sum_{r=0}^{n} V_{n+1}^{(r)}(x) \int_{-\infty}^{\infty} w_r(y)\, dP(y), \tag{4·8}$$

we shall have $P^*(x + \frac{1}{2}) = P_*(x + \frac{1}{2})$ *at all the points* (2·1).

We observe, first, that $V_n(x)$ defines† a symmetrical distribution with finite range $(-\frac{1}{2}n, \frac{1}{2}n)$, mean 0, variance $\frac{1}{12}n$, and $\beta_2 - 3 = -6/(5n)$; it approximates more and more closely to a normal c.d.f. as n increases; and its first $n-1$ derivatives are continuous and vanish at $x = \pm\frac{1}{2}n$, while $V_n^{(n)}(x)$ is a step-function of oscillating sign. Thus, in particular, the

† For the properties of $V_n(x)$ cf. Uspensky (1937, pp. 277, 278 and 305), Kendall (1943, pp. 242 and 245), and Haldane (1945). An explicit formula for $V_n(x)$ is quoted below, § 5 (5·9).

expansion in (4·3), or (4·4), can only represent a true c.d.f. when $a_n = 0$. Secondly, in (4·6), or (4·7), the range is $(-\frac{1}{2}n-\frac{1}{2}, \frac{1}{2}n+\frac{1}{2})$, and the continuity properties of these expansions depend on those of the arbitrary component in W_{n+1}. The sum in (4·7), or (4·8), is always continuous. Thirdly, it may be noted that the polynomials $w_r(y)$ and $\overline{w}_r(y)$ introduced in (4·5) are expressible in terms of generalized Bernoulli polynomials† as follows:

$$w_r(y) = (-)^r B_r^{(n+1)}(\tfrac{1}{2}n+\tfrac{1}{2}+y)/r!, \quad \overline{w}_r(y) = (-)^r B_r^{(n)}(\tfrac{1}{2}n+y)/r!. \tag{4·9}$$

[*Added* 12 *June* 1950. If we indicate dependence on n by writing $P^*(x,n), a_r(n)$, etc., for $P^*(x)$, a_r, etc., we see from (4·9), (4·4) and (4·8) that

$$w_r(y,n) = \overline{w}_r(y,n+1), \quad P_*(x,n) = P^*(x,n+1) - a_{n+1}(n+1)\, V_{n+1}^{(n+1)}(x). \tag{4·10}$$

Thus, when $\mu_1', \ldots, \mu_n'$ are known $P_*(x,n)$ may be calculated as follows: determine a fictitious value μ_{n+1}^*, as a substitute for the possibly unknown moment μ_{n+1}', in such a way as to reduce to zero the coefficient $a_{n+1}(n+1)$, which by (4·9) equals the expectation of

$$(-)^{n+1} B_{n+1}^{(n+1)}(\tfrac{1}{2}n+\tfrac{1}{2}+y)/(n+1)!;$$

then calculate $P^*(x,n+1)$ as if the first $n+1$ moments were $\mu_1', \ldots, \mu_n', \mu_{n+1}^*$.]

5. Proof of (I)

To prove (I) it will suffice to establish the expansion

$$K_n(x,y) = \sum_{r=0}^{n} v_r(x)\, w_r(y) \tag{5·1}$$

with $w_r(y)$ given by (4·5) and $v_r(x)$ by

$$v_r(x) = V_{n+1}^{(r+1)}(x) \quad (r = 0, 1, \ldots, n). \tag{5·2}$$

In fact, substitution of (5·1) in (2·12) gives, as in (3·6),

$$\overline{p}^*(x) = \sum_{r=0}^{n} a_r v_r(x) = \sum_{r=0}^{n} a_r V_{n+1}^{(r+1)}(x), \tag{5·3}$$

$$\begin{aligned} a_r = \int_{-\infty}^{\infty} w_r(y)\, \overline{p}(y)\, dy &= \int_{-\infty}^{\infty} w_r(y)\, dy \int_{y-\frac{1}{2}}^{y+\frac{1}{2}} dP(t) \\ &= \int_{-\infty}^{\infty} dP(t) \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} w_r(y)\, dy = \int_{-\infty}^{\infty} \overline{w}_r(t)\, dP(t). \end{aligned} \tag{5·4}$$

Now, from (2·11), $\delta P^*(x) = \overline{p}^*(x)$ and, from the definition of $V_n(x)$ and $V_{n+1}(x)$,

$$V_{n+1}^{(r+1)}(x) = D^r V_{n+1}'(x) = D^r \int_{-\frac{1}{2}}^{\frac{1}{2}} V_n'(x-t)\, dt = D^r \delta V_n(x) = \delta V_n^{(r)}(x).$$

Hence the relation (5·3), with (5·4), is equivalent to

$$\delta P^*(x) = \delta \sum_{r=0}^{n} a_r V_n^{(r)}(x), \tag{5·5}$$

and thus to (4·4) since both sides of the latter vanish for $x < -\frac{1}{2}n - \frac{1}{2}$. Similarly, the condition (4·3), for (4·2) to hold identically, is equivalent to a special case of (3·9) (with $\overline{p}(x)$ for $g(x)$, $v_r(x)$ for $\alpha_r(x)$, and $w_r(y)$ for $\beta_r(y)$). Lastly, the sum in (4·3), or (4·4), can only have a continuous derivative of order $n-1$ (and therefore also of all lower orders) if all the coefficients vanish except a_0; for the kth derivative of $a_r V_n^{(r)}(x)$ is continuous for $k < n-r$, is discontinuous unless $a_r = 0$ for $k = n-r$, and vanishes except at isolated points for $k > n-r$.

† For relevant formulae see Milne-Thomson (1933, §§6·1–6·4, pp. 126–30, Ex. 1 (ii), p. 150, and §7·03, especially p. 160).

To obtain the expansion (5·1) we shall first resolve $K_n(x,y)$ into a sum of a variable number of terms, each of which enters the sum as (x,y) crosses one of the lines of discontinuity $x-\xi+\frac{1}{2}=0$ from left to right, and shall then expand each such term in descending powers of $x-\xi+\frac{1}{2}$; the effect will be to analyse $K_n(x,y)$ into components of which the first has a continuous $(n-1)$th derivative (with respect to x) and the last is itself discontinuous.

For the first step we require the identity

$$l_\xi(\xi+t)=\sum_{i=0}^{\xi+\frac{1}{2}n}(-)^i\binom{n+1}{i}l_{-\frac{1}{2}n}(i-\tfrac{1}{2}n+t)\quad(\xi+\tfrac{1}{2}n=0,1,\ldots,n). \tag{5·6}$$

Since l_ξ and $l_{-\frac{1}{2}n}$ are polynomials of degree n in t it will suffice to establish (5·6) for the $n+1$ values $t=-\frac{1}{2}n-\xi,\ -\frac{1}{2}n-\xi+1,\ldots,\frac{1}{2}n-\xi$. But for these (i) when $1\leqslant t\leqslant\frac{1}{2}n-\xi$ (in the cases $\xi\leqslant\frac{1}{2}n-1$) all the terms in (5·6) vanish, (ii) when $t=0$ the only non-vanishing terms are $l_\xi(\xi)=1$ on the left and $l_{-\frac{1}{2}n}(-\frac{1}{2}n)=1$ on the right, (iii) when $-\frac{1}{2}n-\xi\leqslant t\leqslant-1$ (in the cases $\xi\geqslant-\frac{1}{2}n+1$) the left-hand side vanishes and the right-hand side is unaltered by the inclusion of the further terms with $i=\xi+\frac{1}{2}n+1,\ldots,n+1$, since these all vanish, and it then becomes $\delta^{n+1}l_{-\frac{1}{2}n}(t+\frac{1}{2})$, which vanishes because $l_{-\frac{1}{2}n}$ is of degree only n.

Secondly, using (2·6) and (2·7), we find

$$\begin{aligned}l_{-\frac{1}{2}n}(\xi+y-x)&=(-)^n\phi_n(y-x+\xi-\tfrac{1}{2})/n!\\&=(-1)^n\sum_{j=0}^{n}\frac{(-x+\xi-\frac{1}{2})^j}{j!}\frac{\phi_n^{(j)}(y)}{n!}\\&=\sum_{r=0}^{n}\frac{(x+\frac{1}{2}-\xi)^{n-r}w_r(y)}{(n-r)!}\end{aligned} \tag{5·7}$$

by application of Taylor's theorem to the polynomial ϕ_n and use of (4·5). On writing $i-\frac{1}{2}n$ for ξ in (5·7) and substituting in (5·6) we get from (2·13), for

$$-\tfrac{1}{2}\leqslant x-\xi<\tfrac{1}{2}\quad\text{and}\quad\xi+\tfrac{1}{2}n=0,1,\ldots,n,$$

$$\begin{aligned}K_n(x,y)&=\sum_{i=0}^{\xi+\frac{1}{2}n}(-)^i\binom{n+1}{i}l_{-\frac{1}{2}n}(i-\tfrac{1}{2}n+y-x)\\&=\sum_{r=0}^{n}w_r(y)\sum_{i=0}^{\xi+\frac{1}{2}n}(-)^i\binom{n+1}{i}\frac{(x+\frac{1}{2}+\frac{1}{2}n-i)^{n-r}}{(n-r)!}\\&=\sum_{r=0}^{n}w_r(y)\,V_{n+1}^{(r+1)}(x),\end{aligned} \tag{5·8}$$

as required for (5·1), in virtue of the known formula†

$$V_m(t-\tfrac{1}{2}m)=\sum_{i=0}^{k}(-)^i\binom{m}{i}\frac{(t-i)^m}{m!}\quad(k\leqslant t<k+1,\ k=0,1,\ldots,m-1). \tag{5·9}$$

The cases where $|\xi|\geqslant\frac{1}{2}n+1$ are trivial, since in them both sides of (5·8) vanish.

6. Proof of (II)

Suppose $Q(x)$ to be given by $$Q(x)=\sum_{r=0}^{n}b_k\,V_n^{(r)}(x). \tag{6·1}$$

Then, by (I), writing $\bar{q}(x)$ for $\delta Q(x)$, we have

$$\bar{q}(x)=\int_{-\infty}^{\infty}K_n(x,y)\,\bar{q}(y)\,dy. \tag{6·2}$$

† Cf. Uspensky (1937, pp. 277 and 278), and also the other references given above in §4.

Now let $W(u)$ be the c.d.f. of an arbitrary distribution on the open interval $(-\frac{1}{2}, \frac{1}{2})$ and let $P(x)$ be derived from $Q(x)$ and $W(x)$ by convolution, so that

$$P(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} Q(x-u)\,dW(u). \tag{6·3}$$

Evidently, (6·1) and (6·3) are together equivalent to (4·6). Now, for

$$\xi = i - \tfrac{1}{2}n \quad (i = \ldots, -1, 0, 1, 2, \ldots)$$

we have from (6·2) and (6·3),

$$\begin{aligned}
\bar{p}(\xi) = \delta P(\xi) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \delta Q(\xi-u)\,dW(u) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \bar{q}(\xi-u)\,dW(u) \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} dW(u) \int_{-\infty}^{\infty} K_n(\xi-u, y)\,\bar{q}(y)\,dy \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} dW(u) \int_{-\infty}^{\infty} K_n(\xi, z)\,\bar{q}(z-u)\,dz \\
&= \int_{-\infty}^{\infty} K_n(\xi, z)\,dz \int_{-\frac{1}{2}}^{\frac{1}{2}} \bar{q}(z-u)\,dW(u) \\
&= \int_{-\infty}^{\infty} K_n(\xi, z)\,\bar{p}(z)\,dz = \bar{p}^*(\xi),
\end{aligned} \tag{6·4}$$

since, by (2·13),

$$K_n(\xi-u, z-u) = l_\xi\{\xi+z-u-(\xi-u)\} = l_\xi\{\xi+z-\xi\} = K_n(\xi, z) \quad (-\tfrac{1}{2} \leqslant u < \tfrac{1}{2}).$$

(It is to be noted that the range of integration in (6·4) is only nominally infinite, and that $K_n(\xi, z)\,\bar{q}(z-u)$ is bounded and, except on isolated lines, is continuous.) From (6·4) we have $\delta P(\xi) = \delta P^*(\xi)$, and hence the required relation $P(\xi+\frac{1}{2}) = P^*(\xi+\frac{1}{2})$ by summation, as after (5·5).

For the second part of the theorem we take $W(u)$ to be the c.d.f. of the rectangular distribution on $(-\frac{1}{2}, \frac{1}{2})$. The assertion involving (4·7) then follows at once from the first part. Further, if we now allow $P(x)$ to be any function satisfying (4·1), $P_*(x)$ as defined in (4·8) will be of the form (4·7), and hence we shall have $P_*^*(\xi+\frac{1}{2}) = P_*(\xi+\frac{1}{2})$ at each ξ of the set (2·1). To complete the proof it thus suffices to show that $P(x)$ *and* $P_*(x)$ *have the same moments* (*of orders not exceeding* n), so that $P^*(x) \equiv P_*^*(x)$. Now for $r = 0, 1, \ldots, n$ we have

$$\int_{-\infty}^{\infty} w_r(y)\,dP_*(y) = \sum_{s=0}^{n} c_s \int_{-\infty}^{\infty} w_r(y)\,V_{n+1}^{(s+1)}(y)\,dy = c_r = \int_{-\infty}^{\infty} w_r(y)\,dP(y), \tag{6·5}$$

from (4·8), by the properties of type (3·5) possessed by the functions $v_r(u) = V_{n+1}^{(r+1)}(u)$ and $w_r(u)$ (cf. (3·4) and (5·1)). Since $w_r(y)$ is a polynomial of degree exactly r the required equality of moments follows from (6·5).

7. Discussion of applications to general distributions

It remains for us to consider the bearing of the results above upon the use of the Hartley-Khamis procedure for distributions in general, and we shall discuss four points that are relevant to practical applications or to further research. The first three points will be considered in the present section and the fourth in § 8.

(*a*) *Alternative methods of obtaining intermediate values.* Of various available methods yielding estimates for $P(x)$ at values of x intermediate between the numbers

$$x = \xi_i + \tfrac{1}{2} = i - \tfrac{1}{2}n + \tfrac{1}{2} \quad (i = \ldots, -1, 0, 1, 2, \ldots)$$

we have so far considered only (i) *the continuously movable grid*, which gives $P^*(x)$, as in (2·10) and (2·11). Another method† is (ii) *the fixed grid supplemented by ordinary interpolation based on some of the values* $P^*(\xi_i+\frac{1}{2})$ $(i=\ldots,-2,-1,0,1,2,\ldots)$: e.g. ($\alpha$) Lagrange interpolation using $(i=-1,0,1,\ldots,n)$; or (β) central-difference interpolation using an even number of values; or (γ) the same but using an odd number. A compromise between (i) and (ii) is method (iii) *use of a few positions of the grid supplemented by ordinary interpolation.*‡ Hartley & Khamis (1947, §7) suggest the use of two, three, or even four equally spaced positions; they do not appear to contemplate practical use of a continuously moving grid. Lastly, we may mention (iv) *use of* $P_*(x)$, *defined in* (4·8), *as an estimate for* $P(x)$. We now remark, first, that methods (i) and (iv) preserve moments of orders not exceeding n (cf. §§3 and 6), while methods (ii) and (iii) in general do not; secondly, while most of these methods give *continuous* estimates for $P(x)$, methods (i), (ii) (γ), and (iii) (γ) give discontinuous estimates in general. The discontinuity of $P^*(x)$ (when $a_n \neq 0$) results from the restriction $-\frac{1}{2} \leqslant \eta < \frac{1}{2}$ imposed upon the 'phase' η in (2·10) in order to make $P^*(x)$ one-valued. We note from (4·4) that the discontinuous term in $P^*(x)$ is $a_n V_n^{(n)}(x)$, and from (5·9) that the jumps of $V_n^{(n)}(x)$ are alternately positive and negative and numerically equal to the binomial coefficients of order n. Thus the jumps of $P^*(x)$ can easily be calculated. As an illustration, in the second test-example worked out by Hartley & Khamis (1947, §7), the jumps would be proportional to $1, -6, 15, -20, 15, -6, 1$, and the magnitude of the greatest jump would be about 0·00083, which is about twice as large as the greatest actual discrepancy between such of the estimates as they calculated and the true values.§

[*Added* 12 *June* 1950. This discontinuous final term in $P^*(x)$ is also of interest as giving some indication of the difference made by bringing in one further moment (though without changing the scale of the grid). In fact, from (4·10),

$$a_{n+1}(n+1)\, V_{n+1}^{(n+1)}(x) = P^*(x, n+1) - P_*(x, n),$$

while by §4 (II), we have $P_*(x,n) = P^*(x,n)$ whenever x has one of the values

$$i - \tfrac{1}{2}n - \tfrac{1}{2} \quad (i = \ldots, -1, 0, 1, \ldots, n, n+1, \ldots).$$

Hence at these points, which include all the discontinuities of $P^*(x, n+1)$, we have

$$P^*(x \pm 0, n+1) - P^*(x, n) = a_{n+1}(n+1)\, V_{n+1}^{(n+1)}(x \pm 0). \qquad (7{\cdot}1)]$$

(*b*) *Choice of origin and scale for the grid.* The expansion (4·4) for $P^*(x)$, or the expansion (4·8) for $P_*(x)$, can be somewhat simplified by a suitable linear transformation of the original variate, designed to make the transformed c.d.f. $P(x)$ agree in mean and variance with the leading term; the two following terms then vanish. (Such a transformation is, of course, very usual in connexion with the Type A series: see §8.) This suggests the following rule: *in the absence of other considerations* (*e.g. knowledge of the range*), *when the Hartley-Khamis procedure* (*with continuously movable grid and* $h = 1$) *is to be used, the mean should be reduced to* 0 *and the variance to* $n/12$ (*at least approximately*); and when the alternative estimate is to be used the reductions should be to 0 and $(n+1)/12$ respectively. (In terms of the original variate we

† The use of a fixed grid and 'standard interpolation' was suggested by Hartley & Khamis (1947, §3).

‡ Note that using $n+1$ positions of the grid and Lagrange interpolation *within each of the intervals* $\xi_i \leqslant x < \xi_{i+1}$ we should get the same results as from method (i), since $P^*(x)$ is a polynomial of degree n in each of these intervals.

§ In this example they used only two positions of the grid, for $\eta = 0$ and $\eta = \frac{1}{2}$; the use of $\eta = -\frac{1}{2}$ instead of $\eta = \frac{1}{2}$ would thus have altered one of their estimates by as much as 0·00083. Actually, $\eta = -\frac{1}{2}$ was the less appropriate owing to considerations of range.

should have $h/\sigma = \sqrt{(12/n)}$ in the former case and $h/\sigma = \sqrt{\{12/(n+1)\}}$ in the latter.) Since $V_{n+1}(x)$ is more nearly normal than $V_n(x)$ the latter rule is likely to give the better results for nearly normal distributions. As an example, for the normal distribution itself, taking $n = 6$, $h = 1$, and using only the central position of the grid, the difference between $\overline{p}^*(x)$ and $\overline{p}(x)$ was less than 0·0008 at each of the points (2·1)† when the variance was $\frac{7}{12}$, but was about 0·018 at $x = 0$ when σ^2 was 1, about 0·0015 when σ^2 was $\frac{6}{12} = \frac{1}{2}$, and about 0·024 when σ^2 was $\frac{1}{4}$.

(*c*) *Possible deductions concerning magnitude of error: first method.* In order to obtain any useful definite bounds for the discrepancy between $P^*(x)$ and $P(x)$ we need, of course, more information about the required distribution than is provided by its first n moments. If such information relates directly to $P(x)$, or to $P'(x)$, then the formulae (2·12) and (2·13) furnish the appropriate starting-point for deductions. For instance, *suppose that the range of the required distribution is contained in* $[-\frac{1}{2}n - \frac{1}{2}, \frac{1}{2}n + \frac{1}{2}]$, *that* $P^{(m)}(x)$ *is continuous in this interval, and that* $P'(x), \ldots, P^{(m-3)}(x)$ *all vanish at its end-points, where* m *is an even integer greater than* 2. Then, using the Euler-Maclaurin sum-formula (cf. Milne-Thomson, 1933, §7·5, (4) and (6), pp. 189, 190), we can obtain‡ for the central differences $\overline{p}^*(x)$ and $\overline{p}(x)$ of $P^*(x)$ and $P(x)$ the relation§

$$\overline{p}^*(x) - \overline{p}(x) = -\frac{(n+2)\,B_m}{m!}\left(\frac{\partial}{\partial y}\right)^m [\{K_n(x,y) - C\}\overline{p}(y)] \qquad (7{\cdot}2)$$

for some y such that $|y| < \frac{1}{2}n + 1$, where B_m is a Bernoulli number, $K_n(x,y)$ is given by (2·13), (2·6) and (2·9), and C is an arbitrary constant. From (7·2) we may deduce a formula given (less explicitly) by Hartley & Khamis‖ for the case where x is one of the numbers $-\frac{1}{2}n, -\frac{1}{2}n+1, \ldots, \frac{1}{2}n$; in fact, we have only to put $C = 1$ and state the right-hand side of (7·2) as a sum of terms arising from those in $K_n(x,y) - C$, considered as a polynomial in y. However, while (7·2) is the formula to which we are naturally led by considering the connexion between the Hartley-Khamis procedure and Sheppard's corrections, it does not appear to give results of value in cases of practical interest; moreover, the mth derivative in (7·2) is an awkward expression for which to obtain bounds.

There are various possible ways of avoiding both the disadvantages just mentioned; I shall content myself here with some indications of one line of attack that is capable (in favourable cases) of establishing bounds for $\overline{p}^*(x) - \overline{p}(x)$ that are not too large to be of interest. The idea is to transfer the factor $K_n(x,y)$, which is oscillatory [cf. (2·6) and (2·13)], from the factor that is repeatedly differentiated (in the integrations by parts that lead to the Euler-Maclaurin formula) to the factor that is repeatedly integrated. *Suppose now that* $P'(x), \ldots, P^{(m)}(x)$ *are*

† I understand that in further work (not yet published) Mr Khamis has obtained a closer fit than the above by using the odd values 5 and 7 of n and choosing the ratio h/σ empirically (with the object of covering the 'essential range'). For $n = 5$ he chose $h/\sigma = 1{\cdot}4$, while the second rule above gives $\sqrt{\frac{12}{5}} = \sqrt{2}$, and for $n = 7$ he chose $h/\sigma = 1{\cdot}2$, while this rule gives $\sqrt{\frac{12}{8}} = 1{\cdot}225$.

‡ For the analogous argument concerning Sheppard's corrections see Kendall (1943, pp. 69 and 70) and also Milne-Thomson (1933).

§ *Note* (12 *June* 1950). The rest of this section, which originally only gave formulae essentially equivalent to (7·2) and (7·18), has been greatly enlarged in order to meet a request from the referee for a more detailed discussion and for numerical examples, with special reference to errors in the tail groups for different values of n. Other changes made in response to this request are the additions at the ends of §4 and §7 (*a*) and an expansion of the formulae (8·6) together with the remarks accompanying them.

‖ 1947 ((23), p. 345, and (10), (11), p. 342); it may be noted that the factor $R+1$ in the last two formulae is presumably a misprint for $R+2$, since the Euler-Maclaurin formula has been applied to $k(r, h, x)$ on the interval $[a - \frac{1}{2}h, b + \frac{1}{2}h]$, whose length is $b - a + h = (R+2)\,h$, not $(R+1)\,h$.

continuous everywhere and vanish outside the interval $[-\frac{1}{2}n-\frac{1}{2}, \frac{1}{2}n+\frac{1}{2}]$, *and that* m *is any positive integer; then we have* [cf. (2·12)]

$$\begin{aligned}\overline{p}^*(x)-\overline{p}(x) &= \int_{-\frac{1}{2}n-1}^{\frac{1}{2}n+1} \overline{p}(y)\, K_n(x,y)\, dy - \overline{p}(x) \\ &= (-)^m \int_{-\frac{1}{2}n-1}^{\frac{1}{2}n+1} \overline{p}^{(m)}(y)\, K_n^{(-m)}(x,y)\, dy, \end{aligned} \tag{7·3}$$

by means of m integrations by parts, where

$$(\partial/\partial y)^m K_n^{(-m)}(x,y) = K_n(x,y) \quad (y \neq x), \tag{7·4}$$

$$\left[\left(\frac{\partial}{\partial y}\right)^{m-1} K_n^{(-m)}(x,y)\right]_{y=x-0}^{y=x+0} = -1, \tag{7·5}$$

the function $K_n^{(-m)}(x,y)$ being determined except for an added arbitrary polynomial of degree less than m by (7·4), (7·5), and (when $m>1$) the further condition that $(\partial/\partial y)^{m-2} K_n^{(-m)}(x,y)$ shall be continuous everywhere. From (7·3) we can get, first, by Schwarz's inequality (cf. Cramér, 1946, §9·5, p. 88),

$$|\overline{p}^*(x)-\overline{p}(x)| \leqslant \left(\int_{-\frac{1}{2}n-1}^{\frac{1}{2}n+1} \{K_n^{(-m)}(x,y)\}^2 dy\right)^{\frac{1}{2}} \left(\int_{-\frac{1}{2}n-1}^{\frac{1}{2}n+1} \{\overline{p}^{(m)}(y)\}^2 dy\right)^{\frac{1}{2}}. \tag{7·6}$$

The first integral on the right can be minimized by choosing the arbitrary terms in $K_n^{(-m)}(x,y)$ so as to make this function orthogonal to $1, y, \ldots, y^{m-1}$ on $[-\frac{1}{2}n-1, \frac{1}{2}n+1]$ (cf. Szegö, 1939, Theorem 3·1·1, p. 37). Secondly, *let us make the additional assumption that* $P^{(m+1)}(x)$ *is bounded and integrable on* $[-\frac{1}{2}n-\frac{1}{2}, \frac{1}{2}n+\frac{1}{2}]$. We shall write $p(x)$ for the density $P'(x)$ and put

$$M_m = \max|p^{(m)}(x)| = \max|P^{(m+1)}(x)| \quad (|x| \leqslant \tfrac{1}{2}n+\tfrac{1}{2}). \tag{7·7}$$

Since $p(x)$ vanishes by hypothesis for $|x| > \frac{1}{2}n+\frac{1}{2}$ we have

$$|\overline{p}^{(m)}(y)| = \left|\int_{y'}^{y''} p^{(m)}(x)\, dx\right| \leqslant M_m \min(1, \tfrac{1}{2}n+1-|y|), \tag{7·8}$$

by (2·2), where $[y', y'']$ is the common part of $[y-\frac{1}{2}, y+\frac{1}{2}]$ and $[-\frac{1}{2}n-\frac{1}{2}, \frac{1}{2}n+\frac{1}{2}]$ and is thus of length $\min(1, \frac{1}{2}n+1-|y|)$. From (7·3) and (7·8) we get, for every x,

$$\begin{aligned}|\overline{p}^*(x)-\overline{p}(x)| &\leqslant M_m \int_{-\frac{1}{2}n-1}^{\frac{1}{2}n+1} |K_n^{(-m)}(x,y)| \min(1, \tfrac{1}{2}n+1-|y|)\, dy \\ &= M_m C_{nm}(x), \end{aligned} \tag{7·9}$$

say. While it is not practicable to minimize the integral in (7·9), yet for a given x we can choose the arbitrary terms in $K_n^{(-m)}(x,y)$ so as to make m of its zeros fall at suitable points in $[-\frac{1}{2}n-1, \frac{1}{2}n+1]$. Such a choice also reduces the labour involved in calculating the integral; even so, the computation is heavier than for the first integral in (7·6). The advantage of (7·9) over (7·6) is that supplementary information about the density $p(x)$ is required by the former only in a comparatively simple form [indeed, in the form familiar in remainder terms of formulae for approximate quadrature of $p(x)$].

As an illustration of the calculation of $C_{nm}(x)$ we take the case $n=6$, $m=4$, $x=0$. We obtain, successively [cf. (2·13)],

$$36K_6(0,y) = 36l_0(y) = (1-y^2)(4-y^2)(9-y^2) = 36-49y^2+14y^4-y^6,$$

$$3780K_6^{(-1)}(0,y) = 3780y-1715y^3+294y^5-15y^7-1890\operatorname{sgn} y,$$

$$181440K_6^{(-4)}(0,y) = 8544y^2+7560y^4-686y^6+42y^8-y^{10}-15120|y|^3.$$

The coefficients of y^3, y^2, y and 1 in the last function were chosen to make it vanish at $y = 0$ (twice) and at $y = \pm 2$; its remaining zeros in $[-4, 4]$ were found to be where $|y| = 1{\cdot}187$, $2{\cdot}289$, $3{\cdot}520$ or $3{\cdot}967$; elementary evaluations then yielded (cf. (7·9))

$$C_{6,4}(0) = 0{\cdot}0268. \tag{7·10}$$

Again, the arbitrary terms in $K_6^{(-4)}(\pm 3, y)$ and $K_4^{(-4)}(\pm 2, y)$ having been chosen so as to make these functions vanish where $y + \frac{1}{2} = 0$ (doubly), -2 or 2, similar calculations yielded

$$C_{6,4}(\pm 3) = 0{\cdot}00130, \quad C_{4,4}(\pm 2) = 0{\cdot}00210. \tag{7·11}$$

These correspond to the extreme tail groups for the cases $n = 6$ and $n = 4$ when the grid is in its central position.

As a simple test case for the effectiveness of (7·9) let us consider the Type II distribution with density

$$p(x) = (315/256)\, a^{-1}(1 - x^2/a^2)^4. \tag{7·12}$$

Then, $P_4(u)$ denoting a Legendre polynomial,

$$|p^{\mathrm{iv}}(x)| = (315/256)\, a^{-5} \cdot 2^4 \cdot 4! \, |P_4(x/a)| \leqslant 945/(2a^5) \quad (|x| \leqslant a), \tag{7·13}$$

since $|P_4(u)| \leqslant 1$ for $|u| \leqslant 1$ (cf. Szegö, 1939, (7·21·1), p. 159). We have also $P_4(\pm 1) = 1$ and thus

$$\left.\begin{aligned} &M_4 = M_4(a) = 945/(2a^5); \\ &M_4(2) = 14{\cdot}766, \quad M_4(2{\cdot}5) = 4{\cdot}838, \quad M_4(3) = 1{\cdot}945, \quad M_4(3{\cdot}5) = 0{\cdot}900. \end{aligned}\right\} \tag{7·14}$$

In a case of this sort Hartley & Khamis (1947, §4) would arrange the scale so as to have $a = \frac{1}{2}(n+1)$; thus, with $n = 6$ and $a = 3{\cdot}5$, (7·9)–(7·14) yield

$$|\overline{p}^*(0) - \overline{p}(0)| \leqslant 0{\cdot}0242, \quad |\overline{p}^*(\pm 3) - \overline{p}(\pm 3)| \leqslant 0{\cdot}00118, \tag{7·15}$$

whereas the actual values of the differences here are $0{\cdot}00205$ and $-0{\cdot}000083$, respectively, and $\overline{p}(0) = 0{\cdot}34217$, $\overline{p}(\pm 3) = 0{\cdot}004534$. For $n = 4$, $a = 2{\cdot}5$, we obtain

$$|\overline{p}^*(\pm 2) - \overline{p}(\pm 2)| \leqslant 0{\cdot}0102, \tag{7·16}$$

whereas actually the difference here is $-0{\cdot}00024$ and $\overline{p}(\pm 2) = 0{\cdot}01960$. As a second test case we may consider the Type I distribution with $p(x) = C(1 - x^2/a^2)^4(1 + x/a)$. It turns out that M_4 is then exactly twice as large as for (7·12), and the bounds in (7·15) and (7·16) would therefore be doubled; however, $\overline{p}(3)$ for $n = 6$, $a = 3{\cdot}5$, and $\overline{p}(2)$ for $n = 4$, $a = 2{\cdot}5$, would be nearly doubled also, though $\overline{p}(-3)$ [or $\overline{p}(-2)$] would be much smaller than before.

The values of $C_{nm}(x)$ in (7·10) and (7·11) correspond to the central position of the grid, but can be used also when the grid is shifted for purposes of interpolation, since we may (equivalently) shift the distribution instead of the grid provided that we start now with a $P(x)$ that is constant outside $[-\frac{1}{2}n, \frac{1}{2}n]$. For the example (7·12) we now take $a = \frac{1}{2}n$ instead of $a = \frac{1}{2}n + \frac{1}{2}$, with a consequent magnification of M_4 [cf. (7·14)]; in this way we can obtain from (7·11)–(7·14), when $n = 6$, $a = 3$,

$$|P^*(x) - P(x)| \leqslant 0{\cdot}0026 \quad (-3 \leqslant x < 2),$$

this interval covering one-sixth of the range and actually containing probability 0·0090, and, when $n = 4$, $a = 2$,

$$|P^*(x) - P(x)| \leqslant 0{\cdot}0310 \quad (-2 \leqslant x < 1),$$

this interval covering a quarter of the range and actually containing probability 0·0492. Presumably, better results could be obtained if values of $C_{nm}(x)$ for non-integral values of $x + \frac{1}{2}n$ were calculated specially, since then we could still take $a = \frac{1}{2}n + \frac{1}{2}$.

(*d*) *Magnitude of error: second method.* We have been supposing, so far, that our additional information about the required distribution was in terms of $P(x)$ or $P'(x)$; however (as will often be the case with sampling distributions), this information may relate to the characteristic

function (ch. f.) $\phi(t)$. Formula (5·1) may then be of use, since we can deduce from it *a formula for* $\phi(t)-\phi^*(t)$, where $\phi^*(t)$ is the ch.f. for $P^*(x)$. In fact, from the definition of $V_n(x)$ [cf. §4, (I)] it follows that the ch.f. corresponding to it is $(\sin\frac{1}{2}t)^n/(\frac{1}{2}t)^n$ (cf. Uspensky, 1937, p. 305); hence that for $V_n^{(r)}(x)$ is $(it)^r(\sin\frac{1}{2}t)^n/(\frac{1}{2}t)^n$, and so, by (4·4),

$$\phi^*(t)=\frac{\sin^n\frac{1}{2}t}{(\frac{1}{2}t)^n}\sum_{r=0}^{n}a_r(it)^r. \tag{7·17}$$

Let us introduce the functions $\eta^*(t)$ and $R_n(t)$ by writing

$$(\tfrac{1}{2}t/\sin\tfrac{1}{2}t)^n\{\phi(t)-\phi^*(t)\}=\eta^*(t)-\sum_{r=0}^{n}a_r(it)^r=R_n(t). \tag{7·18}$$

Since (cf. §3) the moments of orders not exceeding n are the same for $P(x)$ and $P^*(x)$, the Maclaurin series for $\phi(t)$ and $\phi^*(t)$ must agree until after the term in t^n; hence $R_n(t)$ *is simply the remainder after the term in* t^n *in the Maclaurin series for* $\eta^*(t)=(\frac{1}{2}t/\sin\frac{1}{2}t)^n\phi(t)$. Similarly, if $\phi_*(t)$ is the ch.f. corresponding to $P_*(x)$, we obtain [cf. (4·8) and (4·10)]

$$(\tfrac{1}{2}t/\sin\tfrac{1}{2}t)^{n+1}\{\phi(t)-\phi_*(t)\}=\eta_*(t)-\sum_{r=0}^{n}c_r(it)^r=r_n(t)=(\tfrac{1}{2}t)^{n+2}\rho_n(t), \tag{7·19}$$

say, where $r_n(t)$ is the remainder after t^n in the Maclaurin series for $\eta_*(t)=(\frac{1}{2}t/\sin\frac{1}{2}t)^{n+1}\phi(t)$. Now [cf. (2·2)] $\overline{p}(x)=\delta P(x)$, and so, if we write $\overline{p}_*(x)=\delta P_*(x)$, we shall have the formula

$$\overline{p}(x)-\overline{p}_*(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\sin\frac{1}{2}t}{\frac{1}{2}t}\mathscr{R}[\{\phi(t)-\phi_*(t)\}e^{-ixt}]\,dt, \tag{7·20}$$

which holds wherever $\overline{p}(x)$ (and therefore also the left-hand side) is continuous, and also a similar formula for $P(x)-P(-x)-\{P_*(x)-P_*(-x)\}$, with $\sin xt$ instead of $(\sin\frac{1}{2}t)e^{-ixt}$ in the integrand (cf. Cramér, 1946, §10·3, p. 93). From (7·19) and (7·20) we may obtain the following *uniform bound* for the left-hand side of the latter:

$$\begin{aligned}|\,\overline{p}(x)-\overline{p}_*(x)\,| &\leqslant\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{\sin\frac{1}{2}t}{\frac{1}{2}t}\{\phi(t)-\phi_*(t)\}\right|dt\\ &\leqslant\frac{1}{2\pi}\int_{-2\pi}^{2\pi}\sin^{n+2}\tfrac{1}{2}t\,|\,\rho_n(t)\,|\,dt+\frac{1}{2\pi}\int_{|t|\geqslant 2\pi}\frac{|\sin\frac{1}{2}t|}{\frac{1}{2}t}\,|\,\phi(t)\,|\,dt\\ &\qquad+\frac{1}{2\pi}\int_{|t|\geqslant 2\pi}\frac{|\sin\frac{1}{2}t|}{\frac{1}{2}t}\,|\,\phi_*(t)\,|\,dt\\ &=I_1+I_2+I_3,\end{aligned} \tag{7·21}$$

say. The range $[-2\pi,2\pi]$ has been separated out because, in general, the series for $\eta_*(t)$ will not converge outside it owing to poles of $\eta_*(t)$ at $t=\pm 2\pi$. Clearly,

$$I_3\leqslant\sum_{r=0}^{n}2^r\,|\,c_r\,|\,S(n+2,n+2-r), \tag{7·22}$$

say, where the c_r's can be calculated from the known moments by (4·8) and (4·5), while $S(k,l)$ is an integral that (when k is even) can be reduced to cosine integrals and so evaluated accurately, but which in most cases can be overestimated with little loss by the formula

$$S(k,l)=\frac{2}{\pi}\int_{\pi}^{\infty}\frac{|\sin\theta|^k\,d\theta}{\theta^l}\leqslant\left(\frac{1}{\pi}\int_0^{\pi}\sin^k\theta\,d\theta\right)\left(\frac{2}{\pi}\int_{\pi}^{\infty}\frac{d\theta}{\theta^l}\right)\quad(k\geqslant 0,l>1). \tag{7·23}$$

This inequality is obtainable by combining the cases $\alpha=0$, $g(\theta)=(j\pi+\theta)^{-l}(j=1,2,\ldots)$ of the inequality

$$\int_{\alpha}^{\pi-\alpha}\sin^k\theta\,g(\theta)\,d\theta\leqslant\left(\frac{1}{\pi-2\alpha}\int_{\alpha}^{\pi-\alpha}\sin^k\theta\,d\theta\right)\left(\int_{\alpha}^{\pi-\alpha}g(\theta)\,d\theta\right), \tag{7·24}$$

which is valid when $0 \leqslant \alpha \leqslant \frac{1}{2}\pi$, $k \geqslant 0$, and $g''(\theta) \geqslant 0$ $(\alpha \leqslant \theta \leqslant \pi - \alpha)$, in virture of Tchebycheff's inequality for integrals (cf. Hardy, Littlewood & Pólya, 1934, Theorem 236, p. 168), since $\sin^k \theta$ and $g(\theta) + g(\pi - \theta)$ are respectively increasing and decreasing for $\alpha \leqslant \theta \leqslant \frac{1}{2}\pi$.

While the methods to be chosen for overestimating the terms I_1 and I_2 in (7·21) must depend upon $\phi(t)$, and in general the complex integral form of the remainder $r_n(t)$ in (7·19) may be needed, a comparatively simple treatment will suffice if $\phi(t)$ is of a suitable special type. For instance, suppose that (i) *the required distribution is symmetrical*, (ii) *the Maclaurin series of the function* $\eta_*(t)$ *in* (7·19) *has all its coefficients non-negative*,† *and* (iii) *this series is convergent for* $|t| \leqslant 2\pi$. We suppose that a convenient explicit formula is available for $\phi(t)$; and also, for definiteness and slightly greater simplicity, that n *is even*. Then, first, $\phi(t) - \phi_*(t)$ is an even function of t and for $|t| \leqslant 2\pi$ is non-negative; further, $\rho_n(-t) = \rho_n(t)$, and, for $0 \leqslant t < 2\pi$, $\rho_n(t) \geqslant 0$, $\rho_n'(t) \geqslant 0$, and $\rho_n''(t) \geqslant 0$. Thus we can overestimate the parts of I_1 by using the inequalities

$$0 \leqslant \rho_n(t) \leqslant \rho_n(\tfrac{1}{2}\pi) \quad (0 \leqslant t \leqslant \tfrac{1}{2}\pi), \qquad 0 \leqslant \tfrac{1}{2}t\,\rho_n(t)\sin^{n+1}\tfrac{1}{2}t \leqslant \phi(t) \quad (\tfrac{3}{2}\pi \leqslant t < 2\pi), \tag{7·25}$$

$$\frac{1}{2\pi}\left(\int_{-\frac{1}{2}\pi}^{-\frac{1}{2}\pi} + \int_{\frac{1}{2}\pi}^{\frac{1}{2}\pi}\right)\sin^{n+2}\tfrac{1}{2}t\,|\,\rho_n(t)\,|\,dt \leqslant \frac{\frac{1}{2}\rho_n(\frac{1}{2}\pi) + \rho_n(\pi) + \frac{1}{2}\rho_n(\frac{3}{2}\pi)}{\pi}\int_{\frac{1}{2}\pi}^{\frac{1}{2}\pi}\sin^{n+2}\theta\,d\theta, \tag{7·26}$$

the latter inequality resulting from (7·24) with $\alpha = \frac{1}{4}\pi$, $k = n+2$, $g(\theta) = \rho_n(2\theta) = \rho_n(t)$, followed by use of the trapezium rule to overestimate the integral of $\rho_n(2\theta)$ concerned. The values of $\rho_n(t)$ on the right-hand sides of (7·25) and (7·26) are obtainable by (7·19) from the corresponding values of $\phi(t)$. The right-hand side of (7·21) can thus be overestimated by a sum of expressions for which bounds are calculable when a formula for $\phi(t)$ is known. The expression for $P(x) - P(-x) - \{P_*(x) - P_*(-x)\}$, mentioned above after (7·20), can be treated on similar lines, though preferably without immediate resort to absolute values in the integrands.

To test the effectiveness of the programme just outlined some numerical results were worked out for the distribution of the sum v of m independent sample values of a variate u distributed with density $\frac{1}{2}\pi\,\mathrm{sech}^2 u$. For v the ch.f. is‡ $\psi(t) = (\frac{1}{2}t/\sinh\frac{1}{2}t)^m$. (It is possible to obtain‡ a formula for the corresponding density, namely (when m is even),

$$(d/dv)^m[\{(\tfrac{1}{2}m-1)^2+v^2\}\{(\tfrac{1}{2}m-2)^2+v^2\}\ldots\{1^2+v^2\}]\,\tfrac{1}{2}v\coth\pi v/\{(m-1)!\},$$

but this would be very awkward for computation except for quite small values of m.) It can be verified easily that if we change the scale so as to make the variance equal to $(n+1)/12$ (according to the second rule put forward in §7 (*b*) above), then instead of $\psi(t)$ we get

$$\phi(t) = \psi\{t(n+1)^{\frac{1}{2}}m^{-\frac{1}{2}}\} = \{\tfrac{1}{2}t(n+1)^{\frac{1}{2}}m^{-\frac{1}{2}}\}^m/\sinh^m\{\tfrac{1}{2}t(n+1)^{\frac{1}{2}}m^{-\frac{1}{2}}\};$$

also that the conditions (i) (ii), and (iii) above are fulfilled if $n+1 \leqslant m$. (Indeed, the condition in the footnote to (ii) is fulfilled.) The numerical results obtained with $m = 10$, $n = 4$, $\sigma^2 = 5/12$, were as follows:

$$|\,\overline{p}(x) - \overline{p}_*(x)\,| \leqslant 0{\cdot}0074 \quad (-\infty < x < \infty), \tag{7·27}$$

$$\left.\begin{aligned} &\overline{p}_*(\tfrac{1}{2}) = 0{\cdot}440190, \quad \overline{p}_*(\tfrac{3}{2}) = 0{\cdot}058464, \quad \overline{p}_*(\tfrac{5}{2}) = 0{\cdot}001346; \\ &|\,P(\tfrac{3}{2}) - P(-\tfrac{3}{2}) - \{P_*(\tfrac{3}{2}) - P_*(-\tfrac{3}{2})\}\,| \leqslant 0{\cdot}0053; \\ &1 - \{P_*(\tfrac{3}{2}) - P_*(-\tfrac{3}{2})\} = 0{\cdot}021069. \end{aligned}\right\} \tag{7·28}$$

† This holds, in particular, when the cumulants $\kappa_2, \kappa_4, \kappa_6, \ldots$ for $P(x)$ exceed, respectively, those for $V_{n+1}(x)$ by numbers that are alternately non-positive and non-negative.

‡ For relevant formulae see Campbell & Foster (1948, formulae 614, p. 70, and 607·0, p. 68).

In the case $m = 10$, $n = 6$, $\sigma^2 = \frac{7}{12}$, the results were

$$|\overline{p}(x) - \overline{p}_*(x)| \leqslant 0{\cdot}0031 \quad (-\infty < x < \infty), \tag{7·29}$$

$$\overline{p}_*(\tfrac{1}{2}) = 0{\cdot}405585, \quad \overline{p}_*(\tfrac{3}{2}) = 0{\cdot}089629, \quad \overline{p}_*(\tfrac{5}{2}) = 0{\cdot}004726, \quad \overline{p}_*(\tfrac{7}{2}) = 0{\cdot}000060;$$

$$\left.\begin{aligned} |P(2) - P(-2) - \{P_*(2) - P_*(-2)\}| &\leqslant 0{\cdot}0020, \\ 1 - \{P_*(2) - P_*(-2)\} &= 0{\cdot}009572. \end{aligned}\right\} \tag{7·30}$$

It may be recalled from §4 (II), that, when n is even, $P^*(j) = P_*(j)$ and $\overline{p}^*(j+\frac{1}{2}) = \overline{p}_*(j+\frac{1}{2})$ for $j = 0, \pm 1, \pm 2, \ldots$, and also, by (4·10), $P_*(x, n) = P^*(x, n+1)$ for a symmetrical distribution.

(*e*) *Magnitude of error: summing-up.* Of the formulae obtained in §7 (*c*) and (*d*), we may say that formula (7·9), which requires knowledge of a bound for $|p^{(m)}(x)|$, provides a means of proving, in suitable cases, that the Hartley-Khamis estimates are at least rough approximations to the true probabilities (though possibly very rough for the extreme tail groups), and that formulae (7·19) and (7·21) can be made to yield, in cases where the characteristic function is known explicitly and is of a suitable special type, a fairly good uniform bound for the (absolute) errors in the (modified) Hartley-Khamis estimates $\overline{p}_*(x)$. Both methods are susceptible of refinement to some extent, and in the numerical examples worked out both give better bounds when the number of moments is increased. Further, in these examples the bounds obtained for absolute errors in the tail groups are smaller than for those in the central group, although the bounds for relative error are larger [cf. (7·10), (7·11), (7·15) and (7·27)–(7·30)]; comparison with the known true values available in the example (7·12) shows that the same relations obtain when we consider actual errors† instead of bounds for errors.

8. Comparison with the Gram-Charlier Type A series

Let us consider a Type A series for $P(x)$ carried as far as the known moments (of orders not exceeding n) permit; we shall write

$$P_G(x) = \sum_{r=0}^{n} d_r U^{(r)}(x), \quad d_r = \int_{-\infty}^{\infty} h_r(y)\, dP(y), \tag{8·1}$$

$$U(x) = \Phi(x/\sigma'), \quad h_r(y) = (-\sigma')^r H_r(y/\sigma')/r! = (-d/dy)^{n-r} h_n(y), \tag{8·2}$$

where $\Phi(u)$ is the reduced normal c.d.f. and $H_r(u)$ the Hermite polynomial of degree r. (For the relevant formulae cf. Cramér (1946, §17·6, pp. 221–3); also Kendall (1943, §§6·20–6·23, pp. 145–8).) Further, in the reduced case where $P(x)$ agrees in mean and variance with $U(x)$ (so that $\mu_1' = 0, \mu_2' = \sigma^2 = \sigma'^2$) we have, on writing $\gamma_{r-2} = \kappa_r/\sigma'^r$ $(r = 3, 4, \ldots)$ for the reduced cumulants,

$$d_0 = 1, \; d_1 = d_2 = 0, \; d_3 = -\gamma_1\sigma^3/3!, \; d_4 = \gamma_2\sigma^4/4!, \; d_5 = -\gamma_3\sigma^5/5!, \; d_6 = (\gamma_4 + 10\gamma_1^2)\sigma^6/6!, \ldots. \tag{8·3}$$

We now note first that there is a formal analogy between the formulae (8·1) and (8·2) for $P_G(x)$ and the formulae (4·8) and (4·5) for $P_*(x)$, with $U(x)$ corresponding to $V_{n+1}(x)$ and $\sigma'^n H_n(y/\sigma')$ to $y^{(n)}$. Moreover, both $P_G(x)$ and $P_*(x)$ are fitted by moments (cf. §6), and both expansions involve a normalized biorthogonal system of functions of the type (3·5) (although in the case of $P_*(x)$ there is no corresponding set of orthogonal polynomials). In the reduced case for both $P_*(x)$ and $P_G(x)$ (where $\mu_1' = 0$ and $(n+1)/12 = \sigma^2 = \sigma'^2$) we have $c_1 = 0 = d_1$; and we find that $w_2(y) \equiv h_2(y)$ and $w_3(y) \equiv h_3(y)$ (so that $c_2 = d_2$ and $c_3 = d_3$), while $w_r(y) - h_r(y)$

† The numerical examples given by Hartley & Khamis (1947, §§4 and 7) also exhibit these relations between actual errors. See also the remarks following (8·6) below.

is of degree only $r-4$ when $4 \leqslant r \leqslant n$; there is also a fairly close mutual approximation between the two leading terms $V_{n+1}(x)$ and $U(x)$ when we take $\sigma'^2 = (n+1)/12$, even for not very large values of n (cf. Uspensky, 1937, p. 305, and Haldane, 1945).

Turning to the expansion of $P^*(x)$ we note that (4·5) yields easily the relation

$$\overline{w}_r(y) = (-d/dy)^{n-r}\overline{w}_n(y). \tag{8·4}$$

We now have an analogy between the formulae (8·1) and (8·2) for $P_G(x)$ and the formulae (4·4) and (8·4) for $P^*(x)$ like that already found for $P_G(x)$ and $P_*(x)$. In the reduced case for both $P_G(x)$ and $P^*(x)$ (where $\mu_1' = 0$ and $n/12 = \sigma^2 = \sigma'^2$) we have $a_1 = 0 = d_1$, $a_2 = d_2$ and $a_3 = d_3$. The mutual approximation between the leading terms, in this case $\Phi\{x(n/12)^{-\frac{1}{2}}\}$ and $V_n(x)$, is, of course, slightly less close than that between $\Phi[x\{(n+1)/12\}^{-\frac{1}{2}}]$ and $V_{n+1}(x)$ noted above.

To illustrate the numerical resemblance between $P_*(x)$ and $P_G(x)$ we consider the case $n = 5$, $\mu_1' = 0$, $\sigma^2 = \sigma'^2 = (5+1)/12 = \frac{1}{2}$; then, by (II),

$$P^*(k) = P_*(k) \quad (k = 0, \pm 1, \pm 2, \pm 3, \ldots).$$

Using these relations and the notations above we find

$$\left.\begin{aligned} P^*(0) - P_G(0) &= 10^{-3}(7{\cdot}564\gamma_1 - 1{\cdot}1347\gamma_3), \\ P^*(1) - P_G(1) &= 10^{-3}(0{\cdot}1774 - 4{\cdot}819\gamma_1 + 1{\cdot}8727\gamma_2 + 0{\cdot}2226\gamma_3), \\ P^*(2) - P_G(2) &= 10^{-3}(-0{\cdot}0918 + 1{\cdot}296\gamma_1 - 0{\cdot}9026\gamma_2 + 0{\cdot}3162\gamma_3), \\ P^*(3) - P_G(3) &= 10^{-3}(0{\cdot}0110 - 0{\cdot}140\gamma_1 + 0{\cdot}1309\gamma_2 - 0{\cdot}0904\gamma_3). \end{aligned}\right\} \tag{8·5}$$

(The results for $x = -1, -2$ and -3 may be obtained from those for 1, 2 and 3 respectively by changing the signs of the terms not containing γ_1 or γ_3.) In particular, if

$$|\gamma_r| \leqslant 0{\cdot}1 \quad (r = 1, 2, 3)$$

these differences are less in absolute value than 0·00087, 0·00087, 0·00035 and 0·00004 respectively; for a symmetrical distribution these figures can be replaced by 0, 0·00037, 0·00019 and 0·00003 respectively. For comparison, when $n = 7, \mu_1' = 0, \sigma^2 = \sigma'^2 = (1{\cdot}2)^{-2}$ (which approximates to $(7+1)/12$), and when also the required distribution is symmetrical, we get

$$\left.\begin{aligned} P^*(0) - P_G(0) &= 0, \\ P^*(1) - P_G(1) &= 10^{-3}(0{\cdot}008 - 0{\cdot}063\gamma_2 - 0{\cdot}298\gamma_4), \\ P^*(2) - P_G(2) &= 10^{-3}(-0{\cdot}001 + 0{\cdot}041\gamma_2 + 0{\cdot}227\gamma_4), \\ P^*(3) - P_G(3) &= 10^{-3}(-0{\cdot}002 - 0{\cdot}062\gamma_2 - 0{\cdot}014\gamma_4), \\ P^*(4) - P_G(4) &= 10^{-3}(0{\cdot}001 + 0{\cdot}016\gamma_2 + 0{\cdot}008\gamma_4). \end{aligned}\right\} \tag{8·6}$$

If $|\gamma_r| \leqslant 0{\cdot}1$ $(r = 2, 4)$ the last four differences are less in absolute value than 0·000044, 0·000028, 0·000010 and 0·000003 respectively. It is clear that by bringing in the sixth order moment we shall usually get a considerably closer agreement between the two estimates. Moreover, both for $n = 5$ and for $n = 7$, the differences are in general smaller in the tails than in the middle, although, of course, the relative discrepancies in the tails may be greater.

The numerical approximation of $P^*(x)$ (or of $P_*(x)$) to $P_G(x)$ (with appropriate σ') appears to be moderately good when $P(x)$ is close to either of the leading terms in the corresponding expansions; however, there are cases, including the first numerical example given by Hartley & Khamis (1947, §4), in which this approximation is less close than that of $P^*(x)$ to $P(x)$. Such cases thus belong to the class of distributions for which the Hartley-Khamis procedure works better than the Type A expansion. It would seem that information about

the extent of this class should be the first objective of further research on the procedure. In the meantime, in view of the formal analogy between the Type A series (8·1) and the expansion (4·4) of $P^*(x)$ we may conjecture that a distribution will belong to this class when its c.d.f. $P(x)$ is closer to $V_n(x)$ than to $\Phi\{x(n/12)^{-\frac{1}{2}}\}$. Owing to the properties of $V_n(x)$ noted in §4 this will often happen when the distribution in question has a finite range and is slightly platykurtic. (The numerical example just quoted is of this type; it is one of Hartley & Khamis's most successful applications of their procedure.) Similar remarks apply to $P_*(x)$.

REFERENCES

CAMPBELL, G. A. & FOSTER, R. M. (1948). *Fourier Integrals for Practical Applications.* New York: Van Nostrand Co.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton: Princeton University Press.

HALDANE, J. B. S. (1945). *Phil. Mag.* **36**, 184–5.

HARDY, G. H., LITTLEWOOD, J. E. & PÓLYA, G. (1934). *Inequalities.* Cambridge University Press.

HARTLEY, H. O. & KHAMIS, S. H. (1947). *Biometrika*, **34**, 340–51.

KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin and Co.

MILNE-THOMSON, L. M. (1933). *The Calculus of Finite Differences.* London: Macmillan.

SZEGÖ, G. (1939). *Orthogonal Polynomials.* New York: American Mathematical Society.

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability.* New York and London: McGraw-Hill Book Co.

ESTIMATION PROBLEMS WHEN A SIMPLE TYPE OF HETEROGENEITY IS PRESENT IN THE SAMPLE

By W. M. LONG

Central Research Establishment, National Coal Board.

Introduction. This paper is concerned with a problem which the writer has often encountered in practical work; the following are three typical examples:

(i) Several operators each make a number of determinations of some physical constant; it is observed that there is an 'operator effect', i.e. that the sets of results obtained by the different operators disagree to an extent which, although small, cannot be dismissed as accidental. It is desired to estimate, and set confidence limits for, the 'true value' of the constant.

(ii) A certain characteristic of the output of a factory is being examined. For various reasons it is impossible to examine the whole of the output, and instead a number of samples are taken on each of several days. A significant day-to-day variation is in evidence; this variation appears to be quite random. Confidence limits are required for the true average value of the characteristic.

(iii) A certain type of house is being erected in varying numbers on several different sites, the time taken to erect each house being observed. It is known that the physical conditions on a site affect the erection times, and that these conditions vary considerably from site to site. It is desired to estimate the average time which would be observed, in the long run, on the average site.

The common feature of these problems is the complexity of the sample, wherein the elements fall into a number of groups, each group being associated with a different set of physical conditions (in example (iii), for instance, each site comprises one set of conditions). The assumption is made that the sets from which the available information is derived constitute a sample from a population of such sets, and the problem is to estimate the mean of the population which is made up from all the possible observations which could arise from all the possible sets.

1. In mathematical terms, we may sum up the situation as follows:

We are given a sample O of values of a variate, O being composed of k subsamples $o_i\,(i = 1, \ldots, k)$; systematic differences of some kind exist between the subsamples. It is required to estimate the mean of the population from which O is drawn.

Any valid solution must depend on what we know, or are prepared to assume, about the parent population from which the observations are drawn, and in particular about the nature of the systematic differences between the subsamples. The assumptions upon which most of the investigation which follows is based are:

(i) That the subsample o_i may be regarded as a random sample from a normal population, p_i say, with mean a_i and variance σ^2. σ^2 is the same for all the subpopulations p_i, but a_i varies. The systematic differences between the subsamples are to be ascribed to the differences between the a_i's.

(ii) That the set of a_i's associated with those p_i which have contributed to O may be regarded as a random sample from a normal population with mean a and variance σ_a^2.

Our problem then is the estimation of a, and we are not interested in the subpopulation means a_i except in so far as they will assist in the solution.

It is not, of course, asserted that these assumptions will always apply to such physical situations as we have described above. They are, however, of a type that will often be made, especially when not much information is available.

Certain relaxations in the assumptions will be considered at a later stage; for the present, however, they will be retained.

2. Denote by x_{ij} the jth element of the subsample o_i; let the number of elements in o_i be n_i and let $\sum_{i=1}^{k} n_i = N$, so that N = the total number of observations in O. Let $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij} \Big/ n_i$ and $\bar{x} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} \Big/ N$.

Consider for the moment a single subsample o_i. The probability of obtaining the observed set of x's is the probability of obtaining a_i on the hypothesis of random sampling from the population $N(a, \sigma_a^2)$* times the conditional probability of obtaining those x's by random sampling from $N(a_i, \sigma^2)$, summed over all the possible a_i's, i.e. is

$$\int_{a_i=-\infty}^{\infty} \{(2\pi)^{\frac{1}{2}(n_i+1)} \sigma^{n_i} \sigma_a\}^{-1} \exp -\frac{1}{2}\left\{\frac{(a_i-a)^2}{\sigma_a^2} + \sum_{j=1}^{n_i} \frac{(x_{ij}-a_i)^2}{\sigma^2}\right\} \prod_{j=1}^{n_i} dx_{ij}\, da_i.$$

Carrying out the necessary integrations of this type we find, after some algebra, that for the whole sample O the probability element is

$$\left\{(2\pi)^N \sigma^{2(N-k)} \prod_{i=1}^{k} n_i(\sigma_a^2 + \sigma^2/n_i)\right\}^{-\frac{1}{2}} \exp -\frac{1}{2} \sum_{i=1}^{k} \left\{\frac{1}{\sigma^2} \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 + \frac{(\bar{x}_i-a)^2}{(\sigma_a^2+\sigma^2/n_i)}\right\} \prod_{i=1}^{k} \prod_{j=1}^{n_i} dx_{ij}.$$

It will be observed that the mean and variance of each subsample are distributed independently of one another and of all the other subsample means and variances; the probability element can obviously be expressed as the product of factors of the required type.

When the n_i's are all equal the estimation problem offers no difficulty. If we put $n_i = n$ for all i, the set of means $\bar{x}_1, \ldots, \bar{x}_k$ may be regarded as a random sample from the population $N(a, \sigma_a^2 + \sigma^2/n)$; *it follows that the statistic*

$$t = \frac{(\bar{x}-a)}{\left\{\sum_{i=1}^{k} (\bar{x}_i-\bar{x})^2/k(k-1)\right\}^{\frac{1}{2}}}$$

is distributed as Student's t with $(k-1)$ *degrees of freedom.* It may readily be verified that this is the solution reached by application of the likelihood ratio technique.

In particular, if $n_i = 1$ for all i, we have a sample of k observations each of which may be regarded as a more or less inaccurate estimate of the corresponding true variate-value a_i, and we have the corollary that *the Student t-test is still exact when the observations on which it is based are inaccurately measured, provided that the errors of measurement have no systematic bias and all follow the same Gaussian law.*

3. We now turn to the case where the n_i's are not all equal.

The maximum likelihood method does not now lead to any practical result. We find for the estimator of a the weighted mean

$$\sum_{i=1}^{k} (\hat{\sigma}_a^2 + \hat{\sigma}^2/n_i)^{-1} \bar{x}_i \Big/ \sum_{i=1}^{k} (\hat{\sigma}_a^2 + \hat{\sigma}^2/n_i)^{-1}.$$

* Here and below this notation is used to indicate a population represented by a normal distribution with mean a and variance σ_a^2.

In numerical cases the weights have to be obtained by the solution of a set of three simultaneous equations, two of which may be of degree up to the $2k$th in the unknowns. We shall not, therefore, consider the maximum likelihood solution any further, beyond noticing that when the n_i's all become large the solution for a becomes $\hat{a} = \sum_{i=1}^{k} \bar{x}_i \Big/ k$, a form to which we shall return later.

The obvious alternative approach is to use some form of weighted mean of the observations, the system of weighting used to be decided by the precision of the estimator obtained and by its adaptability to interval estimation.

On intuitive grounds, there seems no point in using different weights for different observations from the same subsample; this means, in effect, that we need consider only weighted means of the subsample means, which simplifies matters.

4. The 'grand mean', $\bar{x} = \sum_{i=1}^{k} n_i \bar{x}_i \Big/ \sum_{i=1}^{k} n_i = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} \Big/ N$, is at first sight a possible alternative estimator. Difficulties appear, however, when its use for interval estimation is considered. Its sampling variance is $\left(\sum_{i=1}^{k} n_i^2 \Big/ N^2 \right) \sigma_a^2 + \sigma^2/N$, which can be estimated from the sample using a linear combination of sums of squares of the form

$$b_1 \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2/(k-1) + b_2 \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2/(N-k),$$

where the b's are functions of the n_i's. b_2 is negative, being equal to

$$\frac{k}{N\left(N^2 - \sum_{i=1}^{k} n_i^2\right)} \left(\frac{N^2}{k} - \sum_{i=1}^{k} n_i^2 \right).$$

Moreover, the sum of squares $\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2$ is not distributed as χ^2.

5. An exact solution of the interval estimation problem, as the above examples indicate, appears to be a matter of some difficulty. There may, indeed, exist some system of weighting the $\bar{x}_i$'s from which one could derive a Studentized deviate whose sampling distribution would not be too unmanageable, but the writer has not discovered one. Instead, an approximate solution is suggested, based on the unweighted mean of the subsample means, which we shall denote by

$$\bar{x}' = \sum_{i=1}^{k} \bar{x}_i \Big/ k.$$

The sampling variance of $\bar{x}'$ can be shown to be

$$\frac{\sigma_a^2}{k} + \frac{\sigma^2}{k^2} \sum_{i=1}^{k} \frac{1}{n_i}.$$

This variance is readily estimated from the sample, since the unweighted 'between-means' variance $\sum_{i=1}^{k} (\bar{x}_i - \bar{x}')^2/(k-1)$ has expectation

$$\sigma_a^2 + \frac{\sigma^2}{k} \sum_{i=1}^{k} \frac{1}{n_i}.$$

The solution is obtained by defining, by analogy with Student's ratio, a statistic

$$t' = \frac{(\bar{x}' - a)}{\left\{\sum_{i=1}^{k} (\bar{x}_i - \bar{x}')^2 \Big/ k(k-1)\right\}^{\frac{1}{2}}},$$

and assuming as an approximation that t' is distributed as Student's t with $(k-1)$ degrees of freedom.

In attempting to justify this procedure, we have to consider first the efficiency of $\bar{x}'$ and secondly the closeness of the approximation.

At first sight $\bar{x}'$ may appear a rather crude estimator of a, since it takes no account of the differences in the subsample sizes. Closer examination, however, reveals some points favouring its use.

First, as we have already noticed, when the n_i's all become large it tends to coincidence with the maximum likelihood estimator. Secondly, although the grand mean $\bar{x}$ may appear a more natural choice on intuitive grounds, $\bar{x}'$ will often be the more efficient of the two. For if we subtract the sampling variance of $\bar{x}'$ from that of $\bar{x}$ we get

$$\sigma_a^2 \left\{ \frac{\sum_{i=1}^{k} n_i^2}{N^2} - \frac{1}{k} \right\} + \sigma^2 \left\{ \frac{1}{N} - \frac{1}{k^2} \sum_{i=1}^{k} \frac{1}{n_i} \right\} = \frac{\sigma_a^2}{N^2} \left\{ \sum_{i=1}^{k} n_i^2 - \frac{N^2}{k} \right\} + \frac{\sigma^2}{k} \left\{ \frac{k}{N} - \frac{1}{k} \sum_{i=1}^{k} \frac{1}{n_i} \right\}.$$

The term in the first bracket is always positive; that in the second is always negative, from the property that the arithmetic mean of a set of positive numbers, not all alike, is greater than their harmonic mean. Which of the two estimators is the more efficient depends therefore on the values of the n_i's and of the ratio σ^2/σ_a^2, in any particular case. For a given set of n_i's, the smaller σ is compared with σ_a^2, the more likely it is that $\bar{x}'$ is the better estimator.

As a numerical illustration a few examples are given in the tables below of the efficiencies of both $\bar{x}'$ and $\bar{x}$ for some sets of n_i's and a range of values of the ratio σ^2/σ_a^2. There is a point of some difficulty here, namely, how, in terms of the variances σ_a^2 and σ^2, efficiency should be defined. The normal method of obtaining the asymptotic or minimum variance for efficient estimators, by finding the Hessian matrix of the likelihood function, may not be valid, since the subsamples vary in size. Nor is it necessarily correct to proceed by finding the values of c_i which minimize the variance of the linear function $\sum_{i=1}^{k} c_i \bar{x}_i \Big/ \sum_{i=1}^{k} c_i$; if we do so, subject to the condition $\sum_{i=1}^{k} c_i = 1$, we find $c_i = (\sigma_a^2 + \sigma^2/n_i)^{-1} \Big/ \sum_{i=1}^{k} (\sigma_a^2 + \sigma^2/n_i)^{-1}$, giving for the minimum variance $\left\{ \sum_{i=1}^{k} (\sigma_a^2 + \sigma^2/n_i)^{-1} \right\}^{-1} = V_0$, say. This is appropriate when the ratio σ^2/σ_a^2 is known, which is not the case here. Evaluation of the Hessian matrix leads to the same value V_0, and we have therefore used it; the comparison of the variances of $\bar{x}'$ and $\bar{x}$ will still be valid, and the values given for the efficiency of $\bar{x}'$ will, if incorrect, presumably be too low.

The abbreviation 'Ef' is to be read 'Efficiency of'.

The efficiencies to be expected with larger sizes of subsample may be judged simply by adjusting the values of the ratio σ^2/σ_a^2 in Tables 1–4; for, writing $\sigma^2/\sigma_a^2 = r$, we have

$$\mathrm{Ef}(\bar{x}') = \left\{ \sum_{i=1}^{k} (\sigma_a^2 + \sigma^2/n_i)^{-1} \right\}^{-1} \left\{ \frac{\sigma_a^2}{k} + \frac{\sigma^2}{k^2} \sum_{i=1}^{k} \frac{1}{n_i} \right\}^{-1} = \left\{ \sum_{i=1}^{k} \left(1 + \frac{r}{n_i} \right)^{-1} \right\}^{-1} \left\{ \frac{1}{k^2} \sum_{i=1}^{k} \left(1 + \frac{r}{n_i} \right) \right\}^{-1},$$

and, similarly, $\mathrm{Ef}(\bar{x}) = \left\{\sum_{i=1}^{k}\left(1+\frac{r}{n_i}\right)^{-1}\right\}^{-1}\left\{\frac{\sum_{i=1}^{k} n_i^2}{N^2}+\frac{r}{N}\right\}^{-1}$, so that multiplication of r and each of the n_i's by the same scalar quantity does not alter the efficiencies.

Table 1. $k = 5$, $n_i = 1, 2, 2, 2, 5$, *respectively*, $i = 1, \ldots, 5$

σ^2/σ_a^2	0	0·1	1	5	10	100	∞
Ef $(\bar{x}')$	1	0·999	0·974	0·887	0·845	0·781	0·772
Ef $(\bar{x})$	0·758	0·774	0·864	0·964	0·986	1·000	1

Table 2. $k = 5$, $n_i = 2, 4, 5, 8, 10$

σ^2/σ_a^2	0	0·1	1	5	10	100	∞
Ef $(\bar{x}')$	1	1·000	0·988	0·915	0·863	0·754	0·734
Ef $(\bar{x})$	0·805	0·812	0·862	0·946	0·974	0·999	1

Table 3. $k = 5$, $n_i = 1, 5, 7, 15, 25$

σ^2/σ_a^2	0	0·1	1	5	10	100	∞
Ef $(\bar{x}')$	1	0·999	0·944	0·720	0·596	0·368	0·325
Ef $(\bar{x})$	0·607	0·621	0·699	0·833	0·898	0·995	1

Table 4. $k = 10$, $n_i = 1, 2, 2, 2, 2, 4, 5, 5, 8, 10$

σ^2/σ_a^2	0	0·1	1	5	10	100	∞
Ef $(\bar{x}')$	1	0·999	0·970	0·843	0·773	0·650	0·629
Ef $(\bar{x})$	0·681	0·695	0·785	0·921	0·964	0·999	1

The importance of the ratio σ^2/σ_a^2 in deciding whether $\bar{x}'$ is a better estimator than $\bar{x}$ might have been expected; for when σ_a^2 rather than σ^2 is the important component of variance we have then, virtually, a random sample of k observations $a_1 \ldots a_k$ from the population $N(a, \sigma_a^2)$ and the mean $\bar{x}$ approximates to $\sum_{i=1}^{k} n_i a_i \Big/ \sum_{i=1}^{k} n_i$, which is less efficient than the unweighted mean $\sum_{i=1}^{k} a_i \Big/ k$. At the other end of the scale when σ_a^2 is negligible compared with σ^2 we have, effectively, N random observations from $N(a, \sigma^2)$, and a similar situation occurs with the roles of the two estimators reversed.

6. It seems clear that under favourable circumstances the use of $\bar{x}'$ can be recommended. We now attempt to justify the statement that t' is approximately distributed as Student's t.

For our present purposes the sample is to be regarded as consisting of observations $\bar{x}_1, \ldots, \bar{x}_k$, each normally distributed about a, with variances respectively $\sigma_1^2 \ldots \sigma_k^2$, where $\sigma_i^2 = \sigma_a^2 + \sigma^2/n_i$ $(i = 1, \ldots, k)$. The probability element associated with it is

$$\{(2\pi)^{\frac{1}{2}k}\sigma_1 \ldots \sigma_k\}^{-1} \exp -\frac{1}{2}\sum_{i=1}^{k}\frac{(\bar{x}_i - a)^2}{\sigma_i^2} d\bar{x}_1 \ldots d\bar{x}_k.$$

The observations $\bar{x}_1 \ldots \bar{x}_k$ may be taken as the co-ordinates of a point, P say, in a k-dimensional Euclidean hyperspace. Let us assume, as we may do without loss of generality, that a is zero. Consider the line OQ through the origin of co-ordinates O which makes equal angles with all the axes, i.e. whose direction cosines are $(1/\sqrt{k}, \ldots, 1/\sqrt{k})$. It may be shown (see, for example, Kendall, 1943) that if we define a statistic $z = \bar{x}' \left\{\sum_{i=1}^{k} (\bar{x}_i - \bar{x}')^2\right\}^{-\frac{1}{2}}$, then

$$z = \cot\phi, \quad \text{where} \quad \phi = \text{angle}\, POQ.$$

Also $z = t'/\sqrt{(k-1)}$. z is constant over the cone obtained by rotating OP about OQ, and its probability distribution will be given by integrating the sample probability throughout the volume between the cones defined by $z + \frac{1}{2}dz$ and $z - \frac{1}{2}dz$.

If we apply the transformation $\bar{x}_i = \sigma_i \xi_i$ $(i = 1, \ldots, k)$, the cone corresponding to any particular value of z is transformed into a ξ-cone, and the probability density at the point P is now proportional to $\exp(-\frac{1}{2}OP'^2)$, where P' is the transform of P. The probability $\{|z| \geqslant |z_0|\}$ is therefore proportional to the integral of $\exp(-\frac{1}{2}OP'^2)$ throughout the volume enclosed by the ξ-cone corresponding to z_0, the Jacobian of the transformation being a constant. Now $\exp(-\frac{1}{2}OP'^2)$ is constant over the surface of the $(k-1)$-dimensional hypersphere with centre O and radius OP', and the volume cut off on this hypersphere by the ξ-cone is equal to $OP'^{(k-1)}$ times the solid angle, ω say, of the ξ-cone. The probability $\{|z| \geqslant |z_0|\}$ is therefore proportional to

$$\omega \int_0^\infty OP'^{(k-1)} \exp(-\tfrac{1}{2}OP'^2)\, d(OP'),$$

i.e. to ω.

The problem of determining the sampling distribution of z now becomes that of determining in terms of z the solid angle of the corresponding ξ-cone.

This problem, while simple in statement, presents heavy difficulties for values of $k > 3$, due largely to the fact that the ξ-cones are not cones of revolution; their sections by hyperplanes at right angles to their principal axes are hyperellipsoids.

When $k = 2$ the cones degenerate into pairs of lines in two dimensions, and we easily find that the required probability element is

$$f(z)\, dz = \frac{\beta}{2\pi}\frac{(z^2+1)\, dz}{z^4 + (\beta^2 - 2)z^2 + 1}, \quad \text{where} \quad \beta = \frac{4\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}.$$

The sampling distribution of t' is in this case obtained by writing t' for z in the above. When $\sigma_1 = \sigma_2$ this reduces, as it should, to the Student distribution for one degree of freedom, viz.

$$\frac{1}{\pi}\frac{dt'}{1 + t'^2}.$$

The case $k = 3$ is rather more complicated. The equation of the z-cone, in the original axes, is

$$(\bar{x}_1^2+\bar{x}_2^2+\bar{x}_3^2)\alpha-2(\bar{x}_1\bar{x}_2+\bar{x}_2\bar{x}_3+\bar{x}_3\bar{x}_1)=0,$$

where

$$\alpha=(3\cos^2\phi-1)=\frac{2z^2-1}{1+z^2}=\frac{t'^2-1}{1+\frac{1}{2}t'^2},$$

and that of the corresponding ξ-cone is therefore

$$(\sigma_1^2\xi_1^2+\sigma_2^2\xi_2^2+\sigma_3^2\xi_3^2)\alpha-2(\sigma_1\sigma_2\xi_1\xi_2+\sigma_2\sigma_3\xi_2\xi_3+\sigma_3\sigma_1\xi_3\xi_1)=0.$$

Referred to the principal axes of the cone this takes the form

$$\lambda_1X^2+\lambda_2Y^2+\lambda_3Z^2=0.$$

From the geometry of such cones we can always arrange matters such that $\lambda_1\geqslant\lambda_2\geqslant 0$ and $\lambda_3\leqslant 0$ and the solid angle of the cone is then equal to

$$4\left\{\frac{\pi}{2}-\left(\frac{W_2}{W_1}\right)^{\frac{1}{2}}\int_0^{\frac{1}{2}\pi}\left(\frac{1-W_1\sin^2\theta}{1-W_2\sin^2\theta}\right)^{\frac{1}{2}}d\theta\right\},$$

where for convenience we have written $W_1=(\lambda_1-\lambda_2)/\lambda_1$ and $W_2=(\lambda_1-\lambda_2)/(\lambda_1-\lambda_3)$. W_1 and W_2 are both less than unity, provided that λ_1 and λ_2 are not both zero (in which case the cone degenerates into a line and the solid angle is zero). Under these conditions the series expansions of $(1-W_1\sin^2\theta)^{\frac{1}{2}}$ and $(1-W_2\sin^2\theta)^{-\frac{1}{2}}$ are uniformly convergent for all values of θ, and we may expand the integrand in a series in $\sin^2\theta$ and integrate term by term. To find the constant of proportionality by which the result must be multiplied in order to convert it into a probability we note that when $P\{|z|\geqslant|z_0|\}=1$ the cone opens out into a plane and the solid angle then becomes 2π. The required constant is therefore $1/(2\pi)$. Making the appropriate correction we find that

$$P\{|z|\geqslant|z_0|\}=1-\left(\frac{W_2}{W_1}\right)^{\frac{1}{2}}\{1-(W_1-W_2)[\tfrac{1}{4}+\tfrac{3}{64}(W_1+3W_2)+\tfrac{5}{256}(W_1^2+2W_1W_2+5W_2^2)$$
$$+\tfrac{35}{16384}(5W_1^3+9W_1^2W_2+15W_1W_2^2+35W_2^3)+\ldots]\}.$$

An alternative form which converges more rapidly for low values of the probability is

$$P\{|z|\geqslant|z_0|\}=1-\left(\frac{W_2}{W_1}\right)^{\frac{1}{2}}\left\{1-\frac{1}{2}\left(\frac{W_1-W_2}{W_2}\right)[(1-W_2)^{-\frac{1}{2}}-1]-\frac{1}{8}\left(\frac{W_1-W_2}{W_2}\right)^2\left[\frac{3W_2-2}{(1-W_2)^{\frac{1}{2}}}+1\right]-\ldots\right\},$$

this being obtained by writing the integral in the form

$$\int_0^{\frac{1}{2}\pi}\left\{1-\frac{(W_1-W_2)\sin^2\theta}{(1-W_2\sin^2\theta)}\right\}^{\frac{1}{2}}d\theta$$

and expanding in a series.

The λ's, and hence the W's, are obtained by solving the discriminating cubic of the ξ-cone, viz.

$$\lambda^3-\alpha(\sigma_1^2+\sigma_2^2+\sigma_3^2)\lambda^2+(\alpha^2-1)(\sigma_1^2\sigma_2^2+\sigma_2^2\sigma_3^2+\sigma_3^2\sigma_1^2)\lambda-(\alpha-2)(\alpha+1)^2\sigma_1^2\sigma_2^2\sigma_3^2=0.$$

The probability obviously depends on the relative and not the absolute magnitudes of the σ_i's. When they are all equal they may be equated to unity without affecting the result and the cubic then has roots $(\alpha+1)$, $(\alpha+1)$ and $(\alpha-2)$. W_1 and W_2 are now both zero and

$$W_2/W_1=\lambda_1/(\lambda_1-\lambda_3)=\tfrac{1}{3}(\alpha+1)=z^2/(1+z^2).$$

The probability now becomes $1-z/(1+z^2)^{\frac{1}{2}}$ and the probability element is therefore

$$f(z)\,dz = \frac{dz}{(1+z^2)^{\frac{1}{2}}}.$$

As may be seen by substituting $t = \sqrt{2}\,z$, this agrees with the Student distribution for two degrees of freedom.

Table 5. $k = 2$, $n_1 = 2$, $n_2 = 5$

σ^2/σ_a^2	0	0·1	1	10	100	∞
P_e for $P_a = 0{\cdot}05$	0·05	0·0500	0·0497	0·0472	0·0455	0·0452
P_e for $P_a = 0{\cdot}01$	0·01	0·0100	0·0099	0·0094	0·0091	0·0090

Table 6. $k = 2$, $n_1 = 1$, $n_2 = 7$

σ^2/σ_a^2	0	0·1	1	10	100	∞
P_e for $P_a = 0{\cdot}05$	0·05	0·0500	0·0481	0·0386	0·0339	0·0332
P_e for $P_a = 0{\cdot}01$	0·01	0·0100	0·0096	0·0077	0·0068	0·0066

Table 7. $k = 3$, $n_1 = 2$, $n_2 = 3$, $n_3 = 4$

σ^2/σ_a^2	0	0·1	1	10	100	∞
P_e for $P_a = 0{\cdot}05$	0·05	0·0500	0·0498	0·0486	0·0481	0·0478
P_e for $P_a = 0{\cdot}01$	0·01	0·0100	0·0100	0·0097	0·0095	0·0094

Table 8. $k = 3$, $n_1 = 2$, $n_2 = 5$, $n_3 = 10$

σ^2/σ_a^2	0	0·1	1	10	100	∞
P_e for $P_a = 0{\cdot}05$	0·05	0·0500	0·0495	0·0446	0·0402	0·0394
P_e for $P_a = 0{\cdot}01$	0·01	0·0100	0·0099	0·0087	0·0077	0·0076

The probability functions found above are, unlike the Student distribution, dependent on the values assumed by the σ_i's. For purposes of comparison, therefore, it has been found necessary to compute a number of probabilities for various sets of values of the σ_i's. It has been found convenient to calculate the exact probabilities corresponding to the approximate ones given by assuming that t' is distributed as Student's t; we denote these quantities by P_e and P_a respectively. Some examples are given in Tables 5–8; in each case the values of the ratio σ^2/σ_a^2 and of the n_i's, on which P_e depends, are given.

The following points may be noted:

(i) The tables will give probabilities for values of the n_i's other than those shown; Table 5, for instance, may be adjusted to apply to the case $n_1 = 20$, $n_2 = 50$ by multiplying the values of σ^2/σ_a^2 by 10.

(ii) As was to be expected, the errors involved in the approximation increase with increasing dispersion among the σ_i's, whether brought about by an increase in σ^2/σ_a^2 or by a widening of the differences between the n_i's.

(iii) The exact probabilities shown are never greater than the approximate ones. This is certainly always true for the case $k = 2$, provided that the value of t' in question is numerically greater than unity. We may show this by considering the cumulative function of the distribution, which is

$$\frac{1}{\pi}\tan^{-1}\left(\frac{4\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2 t'^2-1}\frac{|t'|}{}\right).$$

Then

$$P_a - P_e = \frac{1}{\pi}\left\{\tan^{-1}\frac{2|t'|}{t'^2-1} - \tan^{-1}\left(\frac{4\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}\frac{|t'|}{t'^2-1}\right)\right\}.$$

The expression in brackets is positive for $|t'|>1$, since $2 \geqslant 4\sigma_1\sigma_2/(\sigma_1^2+\sigma_2^2)$, and the result follows.

Judging by the similarity of the pattern of the results for the case $k = 3$, a similar property would appear to hold there also; although no proof has been found, a few calculations with widely different values in the set of three σ_i's which have been made confirm this conclusion.

It appears that any test based on the approximation will be a conservative one, in these two cases at least.

The closeness of the approximation is somewhat remarkable. There were, of course, *a priori* grounds for expecting good results when σ^2 is small compared with σ_a^2, for then each $\bar{x}_i$ is approximately equal to its subpopulation mean a_i, and we have, virtually, a sample of k values from the normal population of a_i's. The above examples show that we can go rather further than this; that for σ^2 small compared with σ_a^2 the approximate percentage points are practically indistinguishable from the exact ones, and that even for σ^2 fairly large compared with σ_a^2 the errors involved are not likely to be serious, from a practical point of view (a stronger objection to the use of the approximation would be the comparative inefficiency of $\bar{x}'$ in such a situation).

To obtain some idea of the behaviour of the approximation for larger values of k, we may consider the moments of t'. Write

$$t' = (\bar{x}'-a)\Big/\left\{\sum_{i=1}^{k}(\bar{x}_i-\bar{x}')^2/k(k-1)\right\}^{\frac{1}{2}}$$

$$= \left(\frac{k-1}{kA}\right)^{\frac{1}{2}}\sum_{i=1}^{k}(\bar{x}_i-a)\left\{1+\frac{\sum_{i=1}^{k}(\bar{x}_i-a)^2-\frac{\left\{\sum_{i=1}^{k}(\bar{x}_i-a)\right\}^2}{k}-A}{A}\right\}^{-\frac{1}{2}},$$

where $A = \mathscr{E}\left\{\sum_{i=1}^{k}(\bar{x}_i-\bar{x}')^2\right\}$ and is equal to $\{(k-1)/k\}\sum_{i=1}^{k}\sigma_i^2$. Expand t', t'^2, etc., in series form and take expectations (the arithmetical labour in this is much reduced by the use of tables

of symmetric functions such as those of David & Kendall (1949)). We find, writing $(r) = \sum_{i=1}^{k} \sigma_i^r$ $(r = 1, 2, \ldots)$,

$$\mu_1'(t') = \mu_3(t') = 0,$$

$$\begin{aligned}
\mu_2(t') = 1 &- 2\left(\frac{k}{k-1}\right)\frac{1}{(2)^2}\left\{(4) - \frac{(2)^2}{k}\right\} + 2\left(\frac{k}{k-1}\right)^2\frac{1}{(2)^3}\left\{4(6) + \left(\frac{k-10}{k}\right)(4)(2) + \frac{5}{k^2}(2)^3\right\} \\
&- 4\left(\frac{k}{k-1}\right)^3\frac{1}{(2)^4}\left\{12(8) + 2\left(\frac{k-15}{k}\right)(6)(2) + 3\left(\frac{k-6}{k}\right)(4)^2 - \frac{3}{k}\left(\frac{k-17}{k}\right)(4)(2)^2 - \frac{17}{k^3}(2)^4\right\} \\
&+ 4\left(\frac{k}{k-1}\right)^4\frac{1}{(2)^5}\left\{96(10) + 12\left(\frac{k-20}{k}\right)(8)(2) + 8\left(\frac{5k-36}{k}\right)(6)(4)\right. \\
&\left. - \frac{8}{k}\left(\frac{2k-51}{k}\right)(6)(2)^2 + 3\left(\frac{k^2-20k+156}{k^2}\right)(4)^2(2) + \frac{2}{k^2}(15k-302)(4)(2)^3 + \frac{151}{k^4}(2)^5\right\} \\
&- 8\left(\frac{k}{k-1}\right)^5\frac{1}{(2)^6}\left\{480(12) + 48\left(\frac{k-25}{k}\right)(10)(2) + 180\left(\frac{k-8}{k}\right)(8)(4) - \frac{60}{k}\left(\frac{k-34}{k}\right)(8)(2)^2\right. \\
&+ 80\left(\frac{k-9}{k}\right)(6)^2 + 20\left(\frac{k^2-25k+234}{k}\right)(6)(4)(2) + \frac{20}{k^2}\left(\frac{5k-151}{k}\right)(6)(2)^3 \\
&+ 15\left(\frac{k^2-12k+60}{k^2}\right)(4)^3 - \frac{15}{k}\left(\frac{k^2-34k+336}{k^2}\right)(4)^2(2)^2 \\
&\left. - \frac{5}{k^3}\left(\frac{34k-823}{k}\right)(4)(2)^4 - \frac{823}{k^5}(2)^6\right\} + \ldots.
\end{aligned}$$

This is a function of the k variances $\sigma_1^2, \ldots, \sigma_k^2$. Treating this set as a sample of σ_i^2's, and writing λ_r for the ratio of the rth moment about the mean to the rth power of the first moment, i.e. setting

$$\lambda_r = \frac{1}{k}\sum_{i=1}^{k}\left(\sigma_i^2 - \frac{\sum_{i=1}^{k}\sigma_i^2}{k}\right)^r \Bigg/ \left(\frac{1}{k}\sum_{i=1}^{k}\sigma_i^2\right)^r,$$

we find that the expression for $\mu_2(t')$ falls into the form

$$\begin{aligned}
&\left\{1 - \frac{2}{\nu}\lambda_2\right\} + \left\{\frac{2}{\nu} + \frac{2}{\nu^2}(4\lambda_3 + (\nu+3)\lambda_2)\right\} - \left\{\frac{8}{\nu^2} + \frac{4}{\nu^3}(12\lambda_4 + 3(\nu-5)\lambda_2^2 + 2(\nu+10)\lambda_3 + 3(3\nu+12)\lambda_2)\right\} \\
&\quad + \left\{\frac{12(\nu+4)}{\nu^3} + \frac{4}{\nu^4}(96\lambda_5 + 8(5\nu-31)\lambda_3\lambda_2 + 12(\nu+21)\lambda_1 + 3(\nu^2+22\nu-111)\lambda_2^2\right. \\
&\qquad\qquad \left. + 24(3\nu+8)\lambda_3 + 2(3\nu^2+53\nu+12)\lambda_2)\right\} \\
&\quad - \left\{\frac{32(5\nu+12)}{\nu^4} + \frac{8}{\nu^5}(480\lambda_6 + 180(\nu-7)\lambda_4\lambda_2 + 80(\nu-8)\lambda_3^2 + 48(\nu+36)\lambda_5\right. \\
&\qquad + 20(\nu^2+37\nu-234)\lambda_3\lambda_2 + 360(\nu+6)\lambda_4 + 90(\nu^2+5\nu-34)\lambda_2^2 \\
&\qquad \left. + 20(\nu^2+38\nu+48)\lambda_3 + 5\nu(19\nu+132)\lambda_2)\right\} + \ldots,
\end{aligned}$$

where $\nu = k - 1$. Thus, to order ν^{-2},

$$\mu_2(t') = 1 + \frac{2}{\nu} + \frac{4}{\nu^2} - \frac{6\lambda_2}{\nu^2} = \frac{\nu}{\nu-2} - \frac{6\lambda_2}{\nu^2}.$$

In a similar way we find for the fourth moment of t', to order ν^{-2},

$$\mu_4(t') = 3 + \frac{18}{\nu} + \frac{84}{\nu^2} - \frac{6\lambda_2}{\nu} - \frac{6(15\lambda_2 - 8\lambda_3 + 6\lambda_2^2)}{\nu^2}$$

$$= \frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{6\lambda_2}{\nu} - \frac{6(15\lambda_2 - 8\lambda_3 + 6\lambda_2^2)}{\nu^2}.$$

Apart from the corrective terms λ_r, which depend on the relative variation among the σ_i's, these are the moments of Student's t with ν degrees of freedom. It appears, therefore, that if the σ_i's are not very different the approximation should, for fair-sized ν, be reasonably good. Indeed, the absence of a corrective term of order ν^{-1} in the second moment suggests that the approximation should improve, on the whole, with increasing ν; for the σ_i's are of bounded variation, since $\sigma_a^2 \leqslant \sigma_i^2 \leqslant \sigma_a^2 + \sigma^2$, and hence the λ's, though not constants, cannot exceed certain limiting values.

In view of the accuracies obtained in the cases $\nu = 1$ and $\nu = 2$ above, it seems clear therefore that the approximation is a reliable one. It is likely to go wrong only if, in addition to fairly large variations among the subsample sizes being present, the variance σ^2 is large compared with σ_a^2. It is worth noticing, however, that only infrequently will situations arise where σ^2 is known to be much larger than σ_a^2; not because they may not exist, but because they are difficult to detect. A large number of observations is required before the usual between- and within-groups analysis of variance will establish the existence of a σ_a^2 which is small compared with σ^2. Here we may also note that if, in a practical situation where this approximation has been used, an error of judgement has been committed, in that σ_a^2 has been taken to exist when in fact it does not, the consequences should not normally be very serious. As may be seen by putting σ_a^2 zero in the above, the estimates used in the test are still valid, and the main effect is a certain loss in efficiency.

7. The approximate solution just described applies to a rather wider class of problems than that defined at the outset. The essential steps in the derivation of the moments of t', and of its sampling distribution in the two cases considered, do not require the variances σ_i^2 to be of the form $(\sigma_a^2 + \sigma^2/n_i)$, where σ_a^2 and σ^2 are fixed. The distribution and moments are appropriate to the case of infinite repetition of a sampling procedure leading to k independent observations, normally distributed about the same mean, with variances respectively $\sigma_1^2, \ldots, \sigma_k^2$. How the observations come to have these variances is a matter of indifference. In the case we have considered, where the observations are subsample means, it is not necessary that the two variance components σ_a^2 and σ^2 should be constant from one subsample to another, or that σ^2 should be constant within any subsample; internal correlations between any number of observations within a subsample could exist. Physical situations where such complications might arise are not hard to imagine (e.g. in the example quoted above, where different operators each make a number of determinations of some physical constant, each operator's determinations might be made in sets of any number under conditions varying from one set to another, with consequent between-set effects).

8. We conclude with an example of the application of the test to an analysis of variance problem.

Suppose we have a $2 \times m$ analysis of variance table, with unequal cell frequencies. Let the cell means be $\bar{x}_{ij}$ respectively, where $i = 1, 2$ and $j = 1, \ldots, m$. Form the differences

$$d_j = \bar{x}_{1j} - \bar{x}_{2j}.$$

Let $\bar{d} = \sum_{j=1}^{m} d_j/m$. Then on the usual assumptions of normality the d_j's are normally distributed with variances respectively σ_i^2 say, and we may therefore perform an approximate test for differences in the i-classification by setting

$$t' = \frac{\bar{d}}{\left\{\sum_{j=1}^{m} (d_j - \bar{d})^2/m(m-1)\right\}^{\frac{1}{2}}},$$

and assuming that t' is distributed as Student's t.

This is equivalent to a test of the i-classification against the *interaction*, and the same result would have been reached by using the F-test in a simple analysis of variance based on the cell means and ignoring the complication of unequal cell frequencies. A little thought will show that the approximation involved should be a close one when a fair-sized interaction exists (since then a variance of the type of σ_a^2 above, fairly large in comparison with the within-cell variance, would be present), or, if the interaction is not very pronounced, when the cell frequencies are not too different.

REFERENCES

DAVID, F. N. & KENDALL, M. G. (1949). *Biometrika*, **36**, 431.
KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**, 239.

SOME TESTS FOR RANDOMNESS IN PLANT POPULATIONS

By MARJORIE THOMAS

1. Introduction

This paper is concerned with a discussion of three tests designed to detect departure from randomness in the distribution of a plant species over an area. Material to which the tests may suitably be applied is obtained by quadrat sampling, as explained in a previous paper (Thomas, 1949). It is assumed that if the plants are randomly distributed, the resulting experimental data may be adequately represented by a Poisson series, and the tests therefore are designed to detect departure from this ideal state in the material collected. One of the main requirements of a test for such a purpose is that it shall be as sensitive as possible, and so the discussion will be largely concerned with the power of the tests under conditions departing from the hypothesis of randomness in a way to be specified.

2. Stevens's test

We begin by considering a test put forward by Stevens (1937) which is concerned with the number of quadrats which contain no plants. If there be N quadrats altogether, enclosing a total of S plants, and if Z of these quadrats be empty, the probability distribution of Z, and hence its mean and standard deviation, can be found on the hypothesis that the plants are randomly distributed. This will be a conditional probability distribution, assuming constant values for S and N; the mean and standard deviation of Z will therefore be functions of these two quantities. For any given sampling data, the statistic

$$\zeta = \frac{\text{Observed value of } Z - \text{Expected value of } Z}{\text{Standard deviation of } Z}$$

may be calculated and referred to the normal scale, which, Stevens has stated, approximately graduates the distribution of Z for large samples. For small samples the exact sampling distribution of Z can be used.

Since $p(Z \mid S, N) = p(Z, S \mid N)/p(S \mid N)$, let us consider first the joint probability distribution of Z and S. The probability of having S plants distributed among N quadrats in such a way that Z quadrats contain no plants is the sum, over all possible partitions of S into $N-Z$ parts, of the probabilities that there shall be $\gamma_1, \gamma_2, \ldots, \gamma_{N-Z}$ plants respectively in $N-Z$ quadrats, and the remaining Z quadrats shall be empty. The γ's are partitions of S, so that

$$\sum_{i=1}^{N-Z} \gamma_i = S, \quad \text{and} \quad \gamma_i > 0 \quad \text{for} \quad i = 1, 2, \ldots, (N-Z).$$

Expressed mathematically, this statement is

$$p(Z, S \mid N) = \sum_{\gamma} \frac{N!}{Z!\,(N-Z)!} P_0^Z \prod_{i=1}^{N-Z} P_{\gamma_i}, \tag{1}$$

where P_k is the probability of observing k plants in any given quadrat, and $\sum_{\gamma}$ denotes summation over all sets of γ's fulfilling the required conditions. We shall assume that the distribution for P_k can be represented by the Double Poisson law

$$P_k = \sum_{r=1}^{k} \frac{e^{-m} m^r}{r!} \frac{e^{-r\lambda} (r\lambda)^{k-r}}{(k-r)!} \quad \text{for} \quad k > 0 \quad \text{and} \quad P_0 = e^{-m}, \tag{2}$$

where $m > 0$ and $\lambda \geqslant 0$ are two parameters.* For this distribution, mean k is $M = m(1+\lambda)$. This series is developed from the assumptions that the plants are distributed in clusters, the number of clusters per quadrat following a Poisson distribution with mean m, and the number of plants per cluster, additional to the first, following a Poisson distribution with mean λ. If $\lambda = 0$, equation (2) reduces to the Poisson law, and the distribution of plants is random without clustering. Thus formally the admissible hypotheses are specified by (2), and we wish to test the composite hypothesis H_0 that $\lambda = 0$, m being unspecified. We are, therefore, concentrating attention upon possible departures from randomness being only towards a clustering tendency, and leaving aside consideration of a too regular distribution.

We may now re-write equation (1) and have

$$p(Z, S \mid N) = \frac{N!}{Z!\,(N-Z)!} e^{-mZ} \sum_{\gamma} \prod_{i=1}^{N-Z} \left[\sum_{r=0}^{\gamma_i} \frac{e^{-m} m^r e^{-r\lambda} (r-\lambda)^{\gamma_i - r}}{r!\,(\gamma_i - r)!} \right], \tag{3}$$

where under the null hypothesis H_0, $\lambda = 0$.

To find $p(S \mid N)$ we sum $p(Z, S \mid N)$ over all possible values of Z. The values which Z may take differ according as $S < N$ or $S \geqslant N$. In the first case, Z may run between $N-S$ and $N-1$, while in the second case the wider range from zero to $N-1$ is possible. We have therefore

$$p_1(S \mid N) = \sum_{Z=N-S}^{N-1} p(Z, S \mid N) \quad \text{for} \quad S < N,$$

$$p_2(S \mid N) = \sum_{Z=0}^{N-1} p(Z, S \mid N) \quad \text{for} \quad S \geqslant N,$$

where suffixes attached to the p's serve to distinguish between the two cases. So we obtain finally

$$\left.\begin{aligned} p_1(Z \mid S, N) &= \frac{p(Z, S \mid N)}{p_1(S \mid N)} \quad \text{for} \quad S < N, \\ p_2(Z \mid S, N) &= \frac{p(Z, S \mid N)}{p_2(S \mid N)} \quad \text{for} \quad S \geqslant N, \end{aligned}\right\} \tag{4}$$

as the distribution of quadrats containing no plants when S and N are known. Equations (3) and (4) have been expressed in a manner which does not lend itself to easy algebraic manipulation. It has not so far been found possible to overcome this difficulty, except in the case $\lambda = 0$. We proceed to consider the behaviour of $p(Z, S \mid N)$ when this is so. It will be convenient to quote here one or two of Stevens's results, which will be found useful a little later on. The notation has been altered slightly from that of the original paper, to suit present purposes.

If $\Delta^{N-Z}O^S$ be the $(N-Z)$th leading difference of the series whose successive terms are Sth powers of the natural numbers, beginning with zero, Stevens shows that

$$\frac{\Delta^{N-Z}O^S}{S!} = \sum_{\gamma} \prod_{i=1}^{N-Z} \frac{1}{\gamma_i!}, \tag{5}$$

and

$$\sum_{Z=0}^{N-1} \frac{N!}{Z!\,(N-Z)!} \Delta^{N-Z}O^S = N^S, \tag{6}$$

where the γ's, as above, are partitions of S into $N-Z$ parts. We note that if $S < N$, some terms of the series in equation (6) are zero, since $\Delta^{N-Z}O^S = 0$ for $N - Z > S$. Stevens further finds

* The simplest way of fitting this series to observed data is by use of the recurrence relation

$$P_k = \sum_{t=1}^{k} \frac{mt e^{-\lambda} \lambda^{t-1}}{k\ (t-1)!} P_{k-t}.$$

the probability distribution of empty quadrats when a total of S plants is distributed at random among N quadrats to be

$$p(Z \mid S, N) = \frac{1}{N^S} \frac{N!}{Z!\,(N-Z)!} \Delta^{N-Z} O^S. \tag{7}$$

From this distribution we can find the mean and variance of Z, namely,

$$\left.\begin{aligned} \mu_1'(Z) &= N\left(1-\frac{1}{N}\right)^S, \\ \mu_2(Z) &= N\left[(N-1)\left(1-\frac{2}{N}\right)^S + \left(1-\frac{1}{N}\right)^S - N\left(1-\frac{1}{N}\right)^{2S}\right]. \end{aligned}\right\} \tag{8}$$

If we now start from a different point, and consider the observed data as a sample drawn from a population which follows a Poisson distribution with mean M, it is easy to show that we arrive at the result given by equation (7), whatever value M may have. For if, in equation (3), we put $\lambda = 0$, $m = M$, we have

$$\begin{aligned} p(Z, S \mid N) &= \frac{N!}{Z!\,(N-Z)!} e^{-MZ} \sum_{\gamma} \prod_{i=1}^{N-Z} \frac{e^{-M} M^{\gamma_i}}{\gamma_i!} \\ &= \frac{N!}{Z!\,(N-Z)!} e^{-NM} M^S \sum_{\gamma} \prod_{i=1}^{N-Z} \frac{1}{\gamma_i!} \\ &= \frac{N!}{Z!\,(N-Z)!} e^{-NM} M^S \frac{\Delta^{N-Z} O^S}{S!} \quad \text{by equation (5).} \end{aligned}$$

Hence

$$\begin{aligned} p(S \mid N) &= e^{-NM} M^S \sum_{Z=0}^{N-1} \frac{N!}{Z!\,(N-Z)!} \frac{\Delta^{N-Z} O^S}{S!} \\ &= \frac{e^{-NM} M^S}{S!} N^S \quad \text{by equation (6).} \end{aligned}$$

If $S < N$, $\sum_{Z=0}^{N-S-1} \frac{N!}{Z!\,(N-Z)!} \frac{\Delta^{N-S} O^S}{S!} = 0$, every term in the summation vanishing separately, so that we finally have

$$p(Z \mid S, N) = \frac{p(Z, S \mid N)}{p(S \mid N)} = \frac{N!}{Z!\,(N-Z)!} \frac{\Delta^{N-Z} O^S}{N^S},$$

a result which agrees with that of Stevens (equation (7)), and which is also seen to be independent of M.

Taking now the Double Poisson series, we may consider the power of Stevens's test with respect to alternative distributions defined by parameters λ and $M = m(1+\lambda)$. If Z_2 be taken as the upper significance level for Z, i.e. is chosen so that $P\{Z \geqslant Z_2 \mid \lambda = 0, N, S\} \leqslant \alpha$ (an expression independent of the unknown M), where α has some convenient value, then the power function of the test is given by

$$P\{Z \geqslant Z_2\} = \sum_{Z=Z_2}^{N-1} p(Z \mid N, S) = \frac{\sum_{Z=Z_2}^{N-1} p(Z, S \mid N)}{p(S \mid N)},$$

since $p(S \mid N)$ is independent of Z. This is a conditional power function within the set of samples with fixed S.

The calculation of the power function is a somewhat laborious task for large values of S and N. To make comparison with other tests, it has however been evaluated in the two cases

(i) $S = 5$, $N = 25$; (ii) $S = 10$, $N = 25$;

taking the level of the probability of rejection, given H_0, at about $\alpha = 0{\cdot}03$ in each case, M being chosen so that the population and sample means were equal, that is, $M = 0{\cdot}2$ for (i) and $M = 0{\cdot}4$ for (ii). It is recognized that this is an artificial situation, and that it would be desirable to obtain some knowledge of the behaviour of the power function when the sample mean differs from its expected value. In particular, we would like to consider the 'overall' power function, i.e.

$$\sum_S p(S \mid N).P\{\text{rejection of } H_0 \mid S, N\}.$$

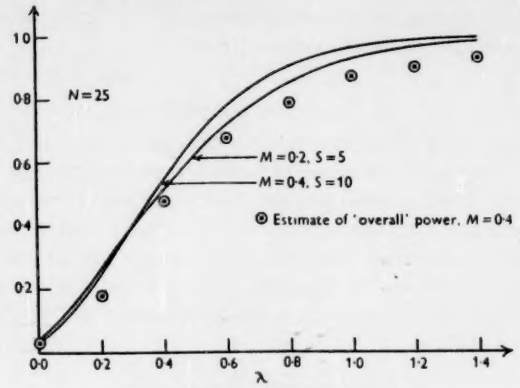


Fig. 1. Power of Stevens's test.

Table 1. *The power of Stevens's test*

λ	Exact 'conditional' power		Estimate of 'overall' power
	$S = 5$	$S = 10$	
	$M = 0{\cdot}2$	$M = 0{\cdot}4$	$M = 0{\cdot}4$
0·0	0·036	0·033	0·030
0·2	0·261	0·251	0·180
0·4	0·515	0·543	0·480
0·6	0·719	0·779	0·680
0·8	0·850	0·905	0·790
1·0	0·925	0·965	0·875
1·2	0·962	0·988	0·900
1·4	0·983	0·997	0·930

Owing to the difficulties presented by discrete distributions, this quantity cannot be handled algebraically, and its exact computation would be very laborious, so estimates of its value for the case $M = 0{\cdot}4$ and various values of λ were found by a sampling experiment. For each value of λ, 200 samples of 25 were drawn from the distribution of equation (2) by means of random sampling numbers (Kendall & Babington Smith, 1939), and Stevens's test applied to each sample. The proportion of samples which led to the rejection of H_0 was taken as an estimate of the overall power of the test. Table 1 shows the results of these calculations, which are also presented graphically in Fig. 1. Comparison in the case of $M = 0{\cdot}4$ of conditional

and overall powers suggests that the power is at or near its optimum when the sample mean has its expected value. Observed differences for the curves for $M = 0{\cdot}2$ and $0{\cdot}4$ suggest also that the power varies considerably with M. This is a drawback, as M will in general be unknown, and so for any given case we shall have little or no idea of the power of the test if the hypothesis of randomness be false, even although the probability of rejecting the hypothesis when it is in fact true is independent of M. For samples as small as 25, we see that the power is not very great*, but we might expect it to be considerably higher for larger samples, say of 100 or 200 quadrats, although the test can never be as powerful as those which take into account the whole plant distribution.

3. Index of dispersion

We will next consider a test for heterogeneity which is based on a comparison of the sample mean and variance. If the null hypothesis, that the plants are randomly distributed, is true, then the population may be described by a simple Poisson distribution, having its mean and variance equal. Alternatively, if there is a tendency to cluster, and we make use of a Double Poisson series to describe observed data, then the variance will be greater than the mean. It seems reasonable, therefore, to take the ratio of sample variance to mean as a measure of heterogeneity. A multiple of this ratio, the index of dispersion, will be denoted by the letter z, so that

$$z = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{\bar{x}},$$

where x_i is the number of plants contained within the ith quadrat, and $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. N is the total number of quadrats. A test suitable for our purposes is simply obtained by choosing a number z_0 and rejecting the hypothesis of random distribution if z, as calculated from the observed sample, exceeds this value. We need therefore to know the probability that, in chance sampling fluctuations, z will exceed z_0, both when the plants are distributed at random and when this is not the case. That is, we need the probability distribution of z. This has long been known to be approximated by a χ^2 distribution with $(N-1)$ degrees of freedom, when the x's are independent variables following a Poisson law. Hoel (1943) has given the first four moments of z in terms of the cumulants of the parent distribution from which the x's are drawn, and confirmed the earlier result. His procedure has been to determine which, if any, of the Pearson type curves, as judged by moments, is applicable as an approximation to the distribution of z, and the method is equally suitable when the parent distribution is not of Poisson form. Bateman (1950) has applied it to the case of a Neyman's contagious population.

The moments of z as used by Hoel are not exact, but approximations derived by the use of Fisher's (1928) k-statistic technique. Following David's (1949) procedure, the first three moments are obtained as follows:

$$\mu_1'(z) = \frac{N-1}{\kappa_1}\left[\kappa_2 + \frac{1}{N}\left\{\frac{\kappa_2^2}{\kappa_1^2} - \frac{\kappa_3}{\kappa_1}\right\} + \frac{1}{N^2}\left\{\frac{3\kappa_2^3}{\kappa_1^4} - \frac{4\kappa_3\kappa_2}{\kappa_1^3} + \frac{\kappa_4}{\kappa_1^2}\right\}\right.$$
$$\left. + \frac{1}{N^3}\left\{\frac{15\kappa_2^4}{\kappa_1^6} - \frac{25\kappa_3\kappa_2^2}{\kappa_1^5} + \frac{7\kappa_4\kappa_2}{\kappa_1^4} + \frac{4\kappa_3^2}{\kappa_1^4} - \frac{\kappa_5}{\kappa_1^3}\right\}\right],$$

* The sampling experiment referred to above has since been extended and it is hoped later to provide a direct comparison of the power of Stevens's test and of that based on the index of dispersion, for the case $N = 25$.

$$\mu_2(z) = \frac{(N-1)^2}{\kappa_1^2}\left[\frac{1}{N}\left\{\frac{\kappa_2^3}{\kappa_1^2} - \frac{2\kappa_3\kappa_2}{\kappa_1} + \kappa_4\right\} + \frac{1}{N-1}2\kappa_2^2 + \frac{1}{N(N-1)}\left\{\frac{6\kappa_2^3}{\kappa_1^2} - \frac{8\kappa_3\kappa_2}{\kappa_1}\right\}\right.$$
$$+ \frac{1}{N^2}\left\{\frac{8\kappa_2^4}{\kappa_1^4} - \frac{18\kappa_3\kappa_2^2}{\kappa_1^3} + \frac{7\kappa_4\kappa_2}{\kappa_1^2} + \frac{5\kappa_3^2}{\kappa_1^2} - \frac{2\kappa_5}{\kappa_1}\right\}$$
$$+ \frac{1}{N^2(N-1)}\left\{\frac{30\kappa_2^4}{\kappa_1^4} - \frac{56\kappa_3\kappa_2^2}{\kappa_1^3} + \frac{12\kappa_4\kappa_2}{\kappa_1^2} + \frac{12\kappa_3^2}{\kappa_1^2}\right\}$$
$$\left. + \frac{1}{N^3}\left\{\frac{69\kappa_2^5}{\kappa_1^6} - \frac{176\kappa_3\kappa_2^3}{\kappa_1^5} + \frac{64\kappa_4\kappa_2^2}{\kappa_1^4} + \frac{84\kappa_3^2\kappa_2}{\kappa_1^4} - \frac{18\kappa_5\kappa_2}{\kappa_1^3} - \frac{26\kappa_4\kappa_3}{\kappa_1^3} + \frac{3\kappa_6}{\kappa_1^2}\right\}\right],$$

$$\mu_3(z) = \frac{(N-1)^3}{\kappa_1^3}\left[\frac{1}{N^2}\left\{\frac{6\kappa_2^5}{\kappa_1^4} - \frac{19\kappa_3\kappa_2^3}{\kappa_1^3} + \frac{9\kappa_4\kappa_2^2}{\kappa_1^2} + \frac{12\kappa_3^2\kappa_2}{\kappa_1^2} - \frac{3\kappa_5\kappa_2}{\kappa_1} - \frac{6\kappa_4\kappa_3}{\kappa_1} + \kappa_6\right\}\right.$$
$$+ \frac{1}{N(N-1)}\left\{\frac{12\kappa_2^4}{\kappa_1^2} - \frac{24\kappa_3\kappa_2^2}{\kappa_1} + 12\kappa_4\kappa_2 + 8\kappa_3^2\right\} + \frac{1}{(N-1)^2}\{8\kappa_2^3 - 4\kappa_3^2\}$$
$$+ \frac{1}{N^3}\left\{\frac{116\kappa_2^6}{\kappa_1^6} - \frac{399\kappa_3\kappa_2^4}{\kappa_1^5} + \frac{168\kappa_4\kappa_2^3}{\kappa_1^4} + \frac{336\kappa_3^2\kappa_2^2}{\kappa_1^4} - \frac{57\kappa_5\kappa_2^2}{\kappa_1^3} - \frac{177\kappa_4\kappa_3\kappa_2}{\kappa_1^3}\right.$$
$$\left. - \frac{44\kappa_3^3}{\kappa_1^3} + \frac{15\kappa_6\kappa_2}{\kappa_1^2} + \frac{30\kappa_5\kappa_3}{\kappa_1^2} + \frac{15\kappa_4^2}{\kappa_1^2} - \frac{3\kappa_7}{\kappa_1}\right\}$$
$$+ \frac{1}{N^2(N-1)}\left\{\frac{144\kappa_2^5}{\kappa_1^4} - \frac{366\kappa_3\kappa_2^3}{\kappa_1^3} + \frac{138\kappa_4\kappa_2^2}{\kappa_1^3} + \frac{204\kappa_3^2\kappa_2}{\kappa_1^2} - \frac{36\kappa_5\kappa_2}{\kappa_1} - \frac{84\kappa_4\kappa_3}{\kappa_1}\right\}$$
$$\left. + \frac{1}{N(N-1)^2}\left\{\frac{48\kappa_2^4}{\kappa_1^2} - \frac{24\kappa_3^2\kappa_2}{\kappa_1^2} + \frac{24\kappa_4\kappa_3}{\kappa_1} - \frac{72\kappa_3\kappa_2^2}{\kappa_1}\right\}\right].$$

Terms within the square brackets of order higher than the third in $1/N$ have been omitted. The κ's are cumulants of the parent population. We shall confine ourselves to consideration of the distribution of z when the parent population is represented by a Double Poisson series, having the following cumulants:

$$\kappa_1 = m(1+\lambda),$$
$$\kappa_2 = m(1+3\lambda+\lambda^2),$$
$$\kappa_3 = m(1+7\lambda+6\lambda^2+\lambda^3),$$
$$\kappa_4 = m(1+15\lambda+25\lambda^2+10\lambda^3+\lambda^4),$$
$$\kappa_5 = m(1+31\lambda+90\lambda^2+65\lambda^3+15\lambda^4+\lambda^5),$$
$$\kappa_6 = m(1+63\lambda+301\lambda^2+350\lambda^3+140\lambda^4+21\lambda^5+\lambda^6),$$
$$\kappa_7 = m(1+127\lambda+966\lambda^2+1701\lambda^3+1050\lambda^4+266\lambda^5+28\lambda^6+\lambda^7).$$

(In the special case $\lambda = 0$, when the series reduces to a simple Poisson, it will be noted that these expressions all become equal to m.)

Numerical values of $\mu_1'(z)$, $\mu_2(z)$ and $\mu_3(z)$ have been computed for $N = 100$ and $N = 200$, and $M = m(1+\lambda) = 1$, 2 and 3 for values of λ ranging from 0·0 to 0·6. Distributions suggesting themselves as suitable to represent that of z are Pearson's Type III and the log-normal distribution, fitted in each case by taking a zero start and equating the first two moments to those of z obtained as above. The ratio $\mu_3(z)$/(third moment of fitted curve) was

calculated for some representative values of M, N and λ, the results indicating that the log-normal was the better choice in this instance, and it was therefore used to obtain the powers shown in Table 2. Powers were also calculated from the fitted Type III curves, and differed very little from those in Table 2, a fact which leads us to hope that the latter are an

Table 2. *Power of the index of dispersion test*

λ	$N = 100$			$N = 200$		
	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$
0·0	0·050	0·050	0·050	0·050	0·050	0·050
0·1	0·345	0·343	0·342	0·535	0·537	0·539
0·2	0·691	0·702	0·705	0·910	0·920	0·923
0·3	0·887	0·900	0·905	0·990	0·993	0·994
0·4	0·965	0·973	0·975	0·999	—	—
0·5	0·990	0·993	0·994	—	—	—
0·6	0·997	—	—	—	—	—
$\log_e z_0$	4·817	4·817	4·817	5·452	5·453	5·453

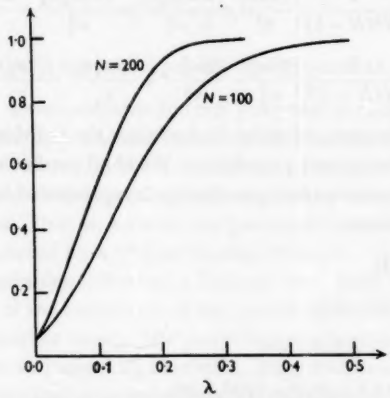


Fig. 2. Power of index of dispersion.

adequate approximation to the true values. We further suggest that, as an alternative to referring $\{\log z - \mathscr{E}(\log z)\}/\sigma(\log z)$ to tables of the unit normal variable (the procedure leading to Table 2), a simpler procedure will be to refer z to a χ^2 distribution with appropriate degrees of freedom. For N greater than 100, the power of the two tests will not differ appreciably. It will be seen that variation in the overall density of the plants, that is, variation of M, has little effect on either the power of the test or the appropriate significance level for the cases considered. Sample size influences the power to some extent, as might have been expected. A graphical comparison is made in Fig. 2, where the values are drawn for $M = 2$. The powers are too close together for all the curves to be drawn, and those for $M = 2$ were selected as being the central ones of those calculated.

4. TEST BASED ON MAXIMUM LIKELIHOOD ESTIMATE OF λ

The last of the tests for randomness which will be considered here is based on a property of maximum likelihood estimates, namely, that such estimates will be consistent and have a sampling distribution which tends to normality for large samples. If l is the maximum likelihood estimate of λ based on the observed frequencies of the zero and unit classes, n_0 and n_1, it was shown in a previous paper (Thomas, 1949) that

$$l = -\log_e\left[-\frac{n_1}{n_0 \log_e (n_0/N)}\right].$$

Further, in the large sample limit, the standard error of l is given by

$$\sigma_l^2 = \frac{m - e^{-m}e^{-\lambda} + e^{-\lambda}(1-m)^2}{Nm^2 e^{-m} e^{-\lambda}},$$

which, when $\lambda = 0$, assumes the value

$$\sigma_l'^2 = \frac{1 - M + M^2 - e^{-M}}{NM^2 e^{-M}}.$$

Hence the power of the test of the hypothesis that $\lambda = 0$, with respect to alternatives $\lambda > 0$ is given by

$$P\{l \geqslant l_\alpha \mid \lambda\} = \frac{1}{\sqrt{(2\pi)}} \int_{(l_\alpha - \lambda)/\sigma_l}^{\infty} e^{-\frac{1}{2}x^2} dx,$$

where $\alpha = \dfrac{1}{\sqrt{(2\pi)}} \displaystyle\int_{l_\alpha/\sigma_l'}^{\infty} e^{-\frac{1}{2}x^2} dx$ is the significance level chosen. Values of this power function are shown below in Table 3 and represented graphically in Fig. 3. A 5 % significance level has

Table 3. *Power of test based on maximum likelihood estimate of* λ

λ	$N = 100$			$N = 200$			$N = 300$		
	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$
0·0	0·050	0·050	0·050	0·050	0·050	0·050	0·050	0·050	0·050
0·1	0·197	0·099	0·058	0·292	0·137	0·074	0·377	0·172	0·087
0·2	0·456	0·197	0·080	0·682	0·324	0·126	0·821	0·439	0·172
0·3	0·711	0·352	0·122	0·916	0·588	0·227	0·978	0·754	0·334
0·4	0·872	0·542	0·193	0·985	0·818	0·387	0·998	0·937	0·563
0·5	0·948	0·720	0·297	0·998	0·942	0·585	—	0·990	0·784
0·6	0·979	0·850	0·431	—	0·986	0·769	—	—	0·924
0·7	0·991	0·927	0·576	—	0·997	0·895	—	—	0·981
0·8	—	0·966	0·710	—	—	0·961	—	—	0·996
1·0	—	0·993	0·891	—	—	0·996	—	—	—
1·2	—	—	0·965	—	—	—	—	—	—
1·4	—	—	0·989	—	—	—	—	—	—

Table 4. *Values of* $l_{0\cdot05}$

M	$N = 100$	$N = 200$	$N = 300$
1	0·216	0·152	0·124
2	0·378	0·268	0·218
3	0·648	0·458	0·374

been taken for all cases. Values of $l_{0\cdot05}$ are shown in Table 4. It will be noticed that the power of the test increases with sample size, which is to be expected. A less desirable variation is that which is observed between different values of plant density for a constant size sample.

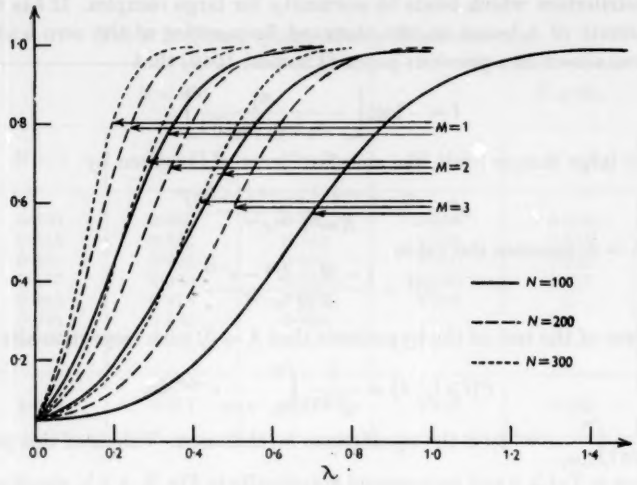


Fig. 3. Power of λ-test.

In practice the value of M is often unknown and must be estimated, and here it is necessary to add a word of caution. If estimates $\hat{M}, l$ of M and λ be made from the same sample, these estimates have a strong positive correlation. If H_0 be true, it is

$$\rho = \frac{M^2 - 2M + 2 - 2e^{-M}}{\sqrt{\{(M^2 - 3M + 4 - 4e^{-M})(M^2 - M + 1 - e^{-M})\}}}$$

which, for example, is 0·846 when $M = 2$. Thus if an estimate $\hat{M}$ of M be used to calculate a value of l_α, the probability that l, when calculated from the same sample, will exceed l_α will very infrequently be even approximately equal to α. For various values of M and N, and $\alpha = 0{\cdot}05$, some values of this probability were worked out, taking $\hat{M}$ in each case as differing from its population value by one standard deviation. Results ranged from $0{\cdot}079\,(M = 1, \hat{M} = M + \sigma_{\hat{M}}, N = 300)$ to a quantity negligible to three decimal places $(M = 3, N = 100, 200, 300)$.

Alternatively, if $\hat{M}$ and l are calculated from different samples, and are therefore independent, an error in $\hat{M}$ of one standard deviation in either direction has less effect. From calculations made, it would appear that for M small ($\leqslant 1$) and N large ($\geqslant 200$), an error of this magnitude in $\hat{M}$ will not affect the significance level when $\alpha = 0{\cdot}05$ by more than 0·01.

Since so little information is needed for the application of this test, we can hardly expect it to be as powerful as others which use the whole range of the sample distribution. In the absence of more complete knowledge, the maximum likelihood method provides a quick, if rough, estimate of the values of M and λ, but when the true value of M is unknown, the suggested test based on these estimates must be used with caution.

5. COMPARISON OF THE TESTS

The third of the three tests described, namely, that based on the maximum likelihood estimate of λ, when only the values n_0, n_1 and N are known, is strictly not comparable with the first two, since it supposes a knowledge of M, the population mean, which they do not. Further, it is only applicable to large samples. Stevens's test and the index of dispersion avoid the first, and Stevens's test the second of these difficulties, but the theoretical treatment of both is awkward, in the one case for large, and in the other for small samples. As far as the present investigation has gone, it appears that the index of dispersion is the statistic to be preferred in testing for randomness in plant populations. The fact that, for large samples, the power of the test varies so little with M is a useful point; in addition, z is easily calculated even without the aid of tables and means of mechanical calculation.

6. SUMMARY

Three tests are described which may be employed to detect departure from randomness in certain plant distributions. The power curves of these tests are plotted under certain specified conditions of departure from the null hypothesis. Some attempt at assessing the relative merits of the tests is made.

I wish to express my grateful thanks to Prof. E. S. Pearson and Dr F. N. David for their help in the preparation of this paper.

REFERENCES

BATEMAN, G. (1950). *Biometrika*, **37**, 59.
DAVID, F. N. (1949). *Biometrika*, **36**, 383.
FISHER, R. A. (1928). *Proc. Lond. Math. Soc.* **30**, 199.
HOEL, P. G. (1943). *Ann. Math. Statist.* **14**, 155.
KENDALL, M. G. & BABINGTON SMITH, B. (1939). *Tracts for Computers*, no. 24.
STEVENS, W. L. (1937). *Ann. Eugen., Lond.*, **8**, 57.
THOMAS, MARJORIE (1949). *Biometrika*, **36**, 18.

CHARTS OF THE POWER FUNCTION FOR ANALYSIS OF VARIANCE TESTS, DERIVED FROM THE NON-CENTRAL *F*-DISTRIBUTION

By E. S. PEARSON and H. O. HARTLEY

1. The distribution of non-central χ^2

If $u_i\,(i = 1, 2, \ldots, \nu)$ form a set of ν random variables each distributed normally and independently with unit variance about a mean zero, and if $a_i\,(i = 1, 2, \ldots, \nu)$ are ν fixed constants, then

$$\chi'^2 = \sum_{i=1}^{\nu} (u_i + a_i)^2 \tag{1}$$

is known to have the following probability density function:

$$p(\chi'^2) = \frac{e^{-\frac{1}{2}\chi'^2} e^{-\frac{1}{2}\lambda}}{2^{\frac{1}{2}\nu}} \sum_{j=0}^{\infty} \frac{(\chi'^2)^{\frac{1}{2}\nu+j-1}\lambda^j}{\Gamma(\frac{1}{2}\nu+j)\,2^{2j}j!}, \tag{2}$$

where

$$\lambda = \sum_{i=1}^{\nu} a_i^2. \tag{3}$$

The function (2) has been termed the distribution of non-central χ^2 having ν degrees of freedom and λ has been termed the non-centrality parameter. The distribution (2) was first obtained by Fisher (1928); in his notation $B^2 = \chi'^2$ and $\beta^2 = \lambda$. Fisher gave a table of the upper 5 % values of B for $\nu = 1\,(1)\,7$, $\beta = 0\,(0{\cdot}2)\,5{\cdot}0$. Subsequently in a Ph.D. Thesis for the University of London (unpublished, 1934), F. Garwood provided a table of the lower 5 % points of B.

If the $a_i = 0$, χ'^2 becomes an ordinary central χ^2. In certain problems it is useful to know, for a given λ, the probability that χ'^2 will exceed a particular percentage point, say the 5 % point, of the ordinary χ^2 distribution. This probability expressed as a function of λ gives the power function of the χ^2 test. Appropriate tables have been computed by Patnaik (1949) and Fix (1949).

2. The distribution of non-central F

If $\chi_1'^2$ is a value of non-central χ^2 having ν_1 degrees of freedom and χ_2^2 a central χ^2 with ν_2 degrees of freedom independent of $\chi_1'^2$, then we may write

$$F' = \frac{\chi_1'^2}{\nu_1} \Big/ \frac{\chi_2^2}{\nu_2} = \frac{\nu_2 \sum_{i=1}^{\nu_1} (u_i + a_i)^2}{\nu_1 \sum_{i=\nu_1+1}^{\nu_1+\nu_2} u_i^2}, \tag{4}$$

where all the $u_i\,(i = 1, 2, \ldots, \nu_1 + \nu_2)$ are independent normal deviates having unit variance and zero mean. F' may be termed a non-central variance ratio. The probability density distribution of F' is known to be

$$p(F') = \sum_{j=0}^{\infty} \left\{ \frac{e^{-\frac{1}{2}\lambda}(\frac{1}{2}\lambda)^j}{j!\,B(\frac{1}{2}\nu_1 + j, \frac{1}{2}\nu_2)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}\nu_1+j} F'^{\frac{1}{2}\nu_1-1+j} \left(1 + \frac{\nu_1}{\nu_2}F'\right)^{-\frac{1}{2}(\nu_1+\nu_2)-j} \right\} \tag{5}$$

where $\lambda = \sum_i a_i^2$ as before. The distribution is immediately derivable from the form given by Fisher (1928) (his distribution C), who took as variable

$$R^2 = \chi_1'^2/(\chi_1'^2 + \chi_2^2).$$

Any general tabulation of the probability integral of (5) would involve extensive computation as three parameters ν_1, ν_2 and λ are involved.

The application of the F'-distribution with which we shall be concerned here is in connexion with the analysis of variance. It is clear that if in a comparison of two variance estimates the underlying variables in the 'treatment' sum of squares are non-central, then the chance of establishing significance for a given non-centrality can be determined from the probability integral of $p(F')$. Thus

$$\int_{F_\alpha}^{\infty} p(F')\,dF' = \beta(\lambda \mid \alpha, \nu_1, \nu_2), \tag{6}$$

where F_α is the 100α % significance level of the ordinary variance ratio test, gives the probability of establishing significance at the 100α % level for a given value of λ. In other words $\beta(\lambda \mid \alpha, \nu_1, \nu_2)$, regarded as a function of λ, is the power function of the analysis of variance test. This use of the distribution of F' has been discussed by Tang (1938), who provided certain tables, and by Patnaik (1949). Also Lehmer (1944) published additional tables and formulae for the asymptotic distribution of F' for large ν_2 and/or ν_1.

In the present paper we have provided charts which, we believe, are easier to handle than these tables. Just as in Tang's tables on which they are largely based, the charts are confined to variance ratio F-tests carried out at the 5 and 1 % levels of significance, i.e. the error of the first kind, α, is confined to values of 0·05 and 0·01. We have also retained Tang's non-centrality parameter

$$\phi = \sqrt{\{\lambda/(\nu_1+1)\}} = \left\{\sum_{i=1}^{\nu_1} a_i^2/(\nu_1+1)\right\}^{\frac{1}{2}}, \tag{7}$$

in place of λ.

Eight charts are provided corresponding to the eight values $\nu_1 = 1\,(1)\,8$. Each chart gives two families of power curves corresponding to $\alpha = 0{\cdot}05$ and 0·01. In each family an individual power curve corresponds to a specified value of ν_2 and gives the power $\beta(\phi \mid \alpha, \nu_1, \nu_2)$ as ordinate plotted against an abscissa scale of ϕ. The curves are drawn for

$$\nu_2 = 6\,(1)\,10, 12, 15, 20, 30, 60 \text{ and } \infty.$$

It should be noted that the β-scale is logarithmic, thereby expanding the scale in the important region of high power $0{\cdot}80 \leqslant \beta \leqslant 0{\cdot}99$, whilst at the same time straightening the curves and facilitating interpolation. The β-grid and the harmonic ν_2-grid are sufficiently fine to permit interpolation by sight. The ϕ-grid is equidistant and the vertical rules correspond to steps of 0·2 (or 0·1) in ϕ.

As we shall see below, in many applications the charts are used to obtain the power, β, corresponding to given values of α, ν_1, ν_2 and ϕ, but often β is specified, and it is required to find ν_2 or ϕ. For such problems where inverse and multivariate interpolation is required, the present graphical form appeared to us to have advantages over tabulation, particularly as high accuracy is unnecessary. In what follows we are concerned with:

(*a*) A description of the method to be adopted in determining the form for ϕ, given the experimental set-up and the effect whose significance is under test.

(*b*) Illustrations of the use of the charts.

3. Application to the analysis of variance

To bring the distribution (5) into play, the mathematical structure which we must regard as appropriate to the experimental data is one in which the random variation is due to a normally distributed error element, superimposed upon a number of additive constant terms. In the most general form, x_i is defined by

$$x_i = \xi_i + z_i \quad (i = 1, 2, \ldots, N), \tag{8}$$

where

(i) $\xi_i = c_{i1}\theta_1 + c_{i2}\theta_2 + \ldots + c_{is}\theta_s \quad (i = 1, 2, \ldots, N)$, (9)

(ii) the z_i are N independent normal deviates each having expectation zero and a common variance σ_z^2.

The θ's are s parameters whose values are generally unknown, while the c's are known constants, usually equalling 1 or 0. The null hypothesis specifies the values of, say, r of the θ's, e.g. $\theta_j = \theta_j^0$ $(j = 1, 2, \ldots, r)$. The power function of the test measures the probability of establishing significance when $\theta_j - \theta_j^0 \neq 0$ $(j = 1, 2, \ldots, r)$, the non-centrality parameter ϕ being a function of these differences and of σ_z.

It is well known that this mathematical model cannot always be regarded as appropriate; for example, the differences $\theta_j - \theta_j^0$, if they are not zero, may be more reasonably regarded as random variables rather than constants. Again, in many applications of analysis of variance the residuals z_i cannot be assumed independent. We think, however, that the model implied by equations (8) and (9) is applicable to a considerable number of problems and hope that by making the present charts available more thought may be focused on the mathematical assumptions underlying methods of practical analysis in common use.*

Starting with this model, Kolodziejczyk (1935) showed how it was possible to determine in a systematic way the two appropriate sums of squares to compare in a test of any hypothesis concerning the values of the θ's. He described the general procedure as the test of a linear hypothesis. In deriving the power function of the test, Tang (1938) also started from this basis. It is usually unnecessary, however, to determine from first principles which are the appropriate sums of squares to compare, as they are known from the well-recognized character of the experimental arrangement. We shall therefore suppose that the two sums of squares appropriate for a given comparison are known; they may be termed the 'treatment' sum of squares S_1 based on ν_1 degrees of freedom and the 'error' sum of squares S_2 based on ν_2 degrees of freedom. All that is necessary for use of the charts is to determine the ϕ of equation (7). As Tang has shown, the procedure is simple:

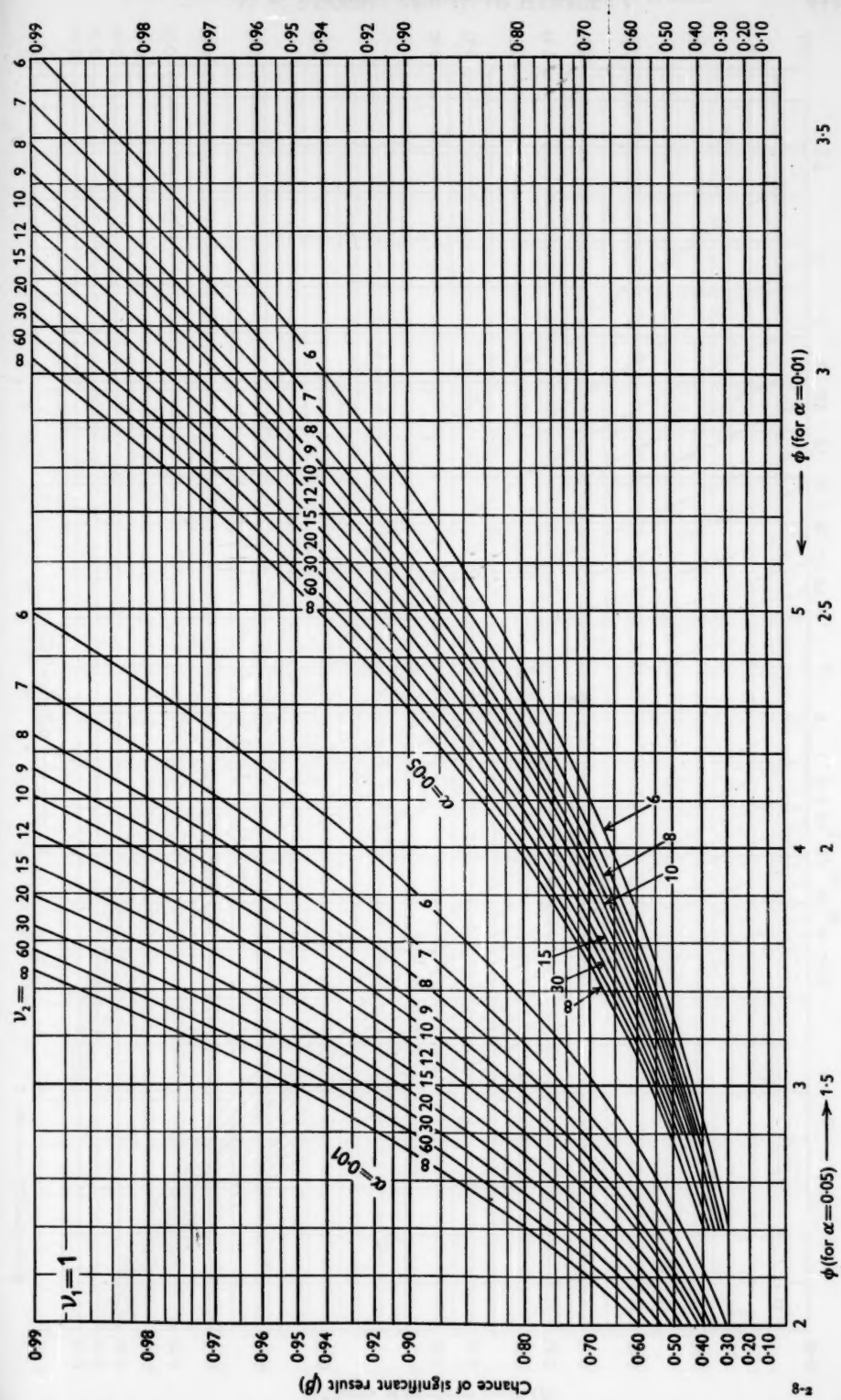
(*a*) We must define the assumed algebraic set-up of equations (8) and (9) and note which are the parameters $\theta_1, \theta_2, \ldots \theta_r$, whose values are specified by the null hypothesis.

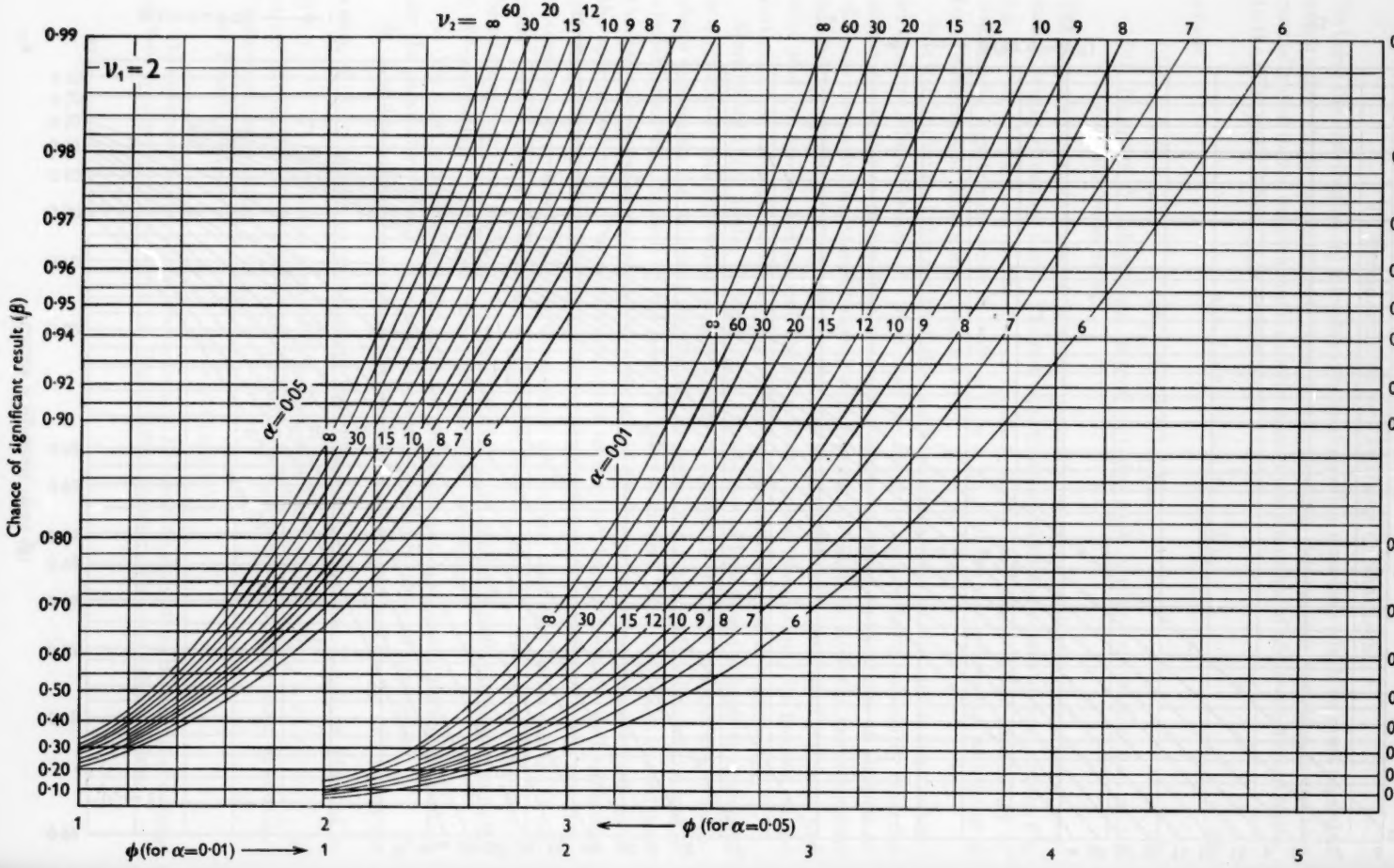
(*b*) S_1 will be expressible as a sum of squares of linear functions of the x_i. $\lambda\sigma_z^2$ is then the value of S_1 obtained by substituting into this sum of squares $\sum_{j=1}^{r} c_{ij}\theta_j$ in place of x_i $(i = 1, 2, \ldots, N)$. Since the form of S_1 is such that the remaining θ_j $(j = r+1, r+2, \ldots, s)$ cancel out on substitution for x_i, the procedure is equivalent to making the z_i in equation (8) zero and substituting ξ_i for x_i.

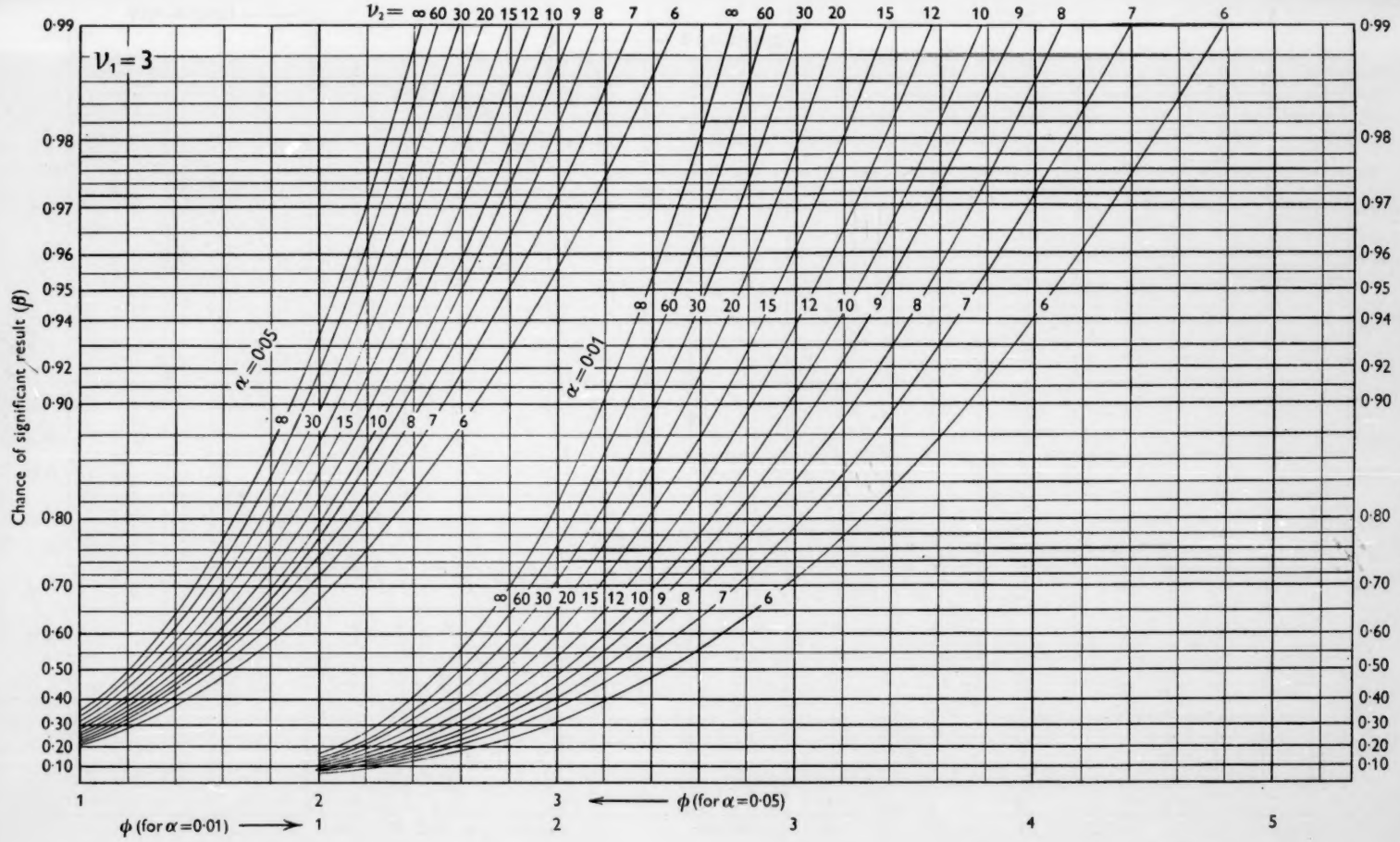
(*c*) Finally we have

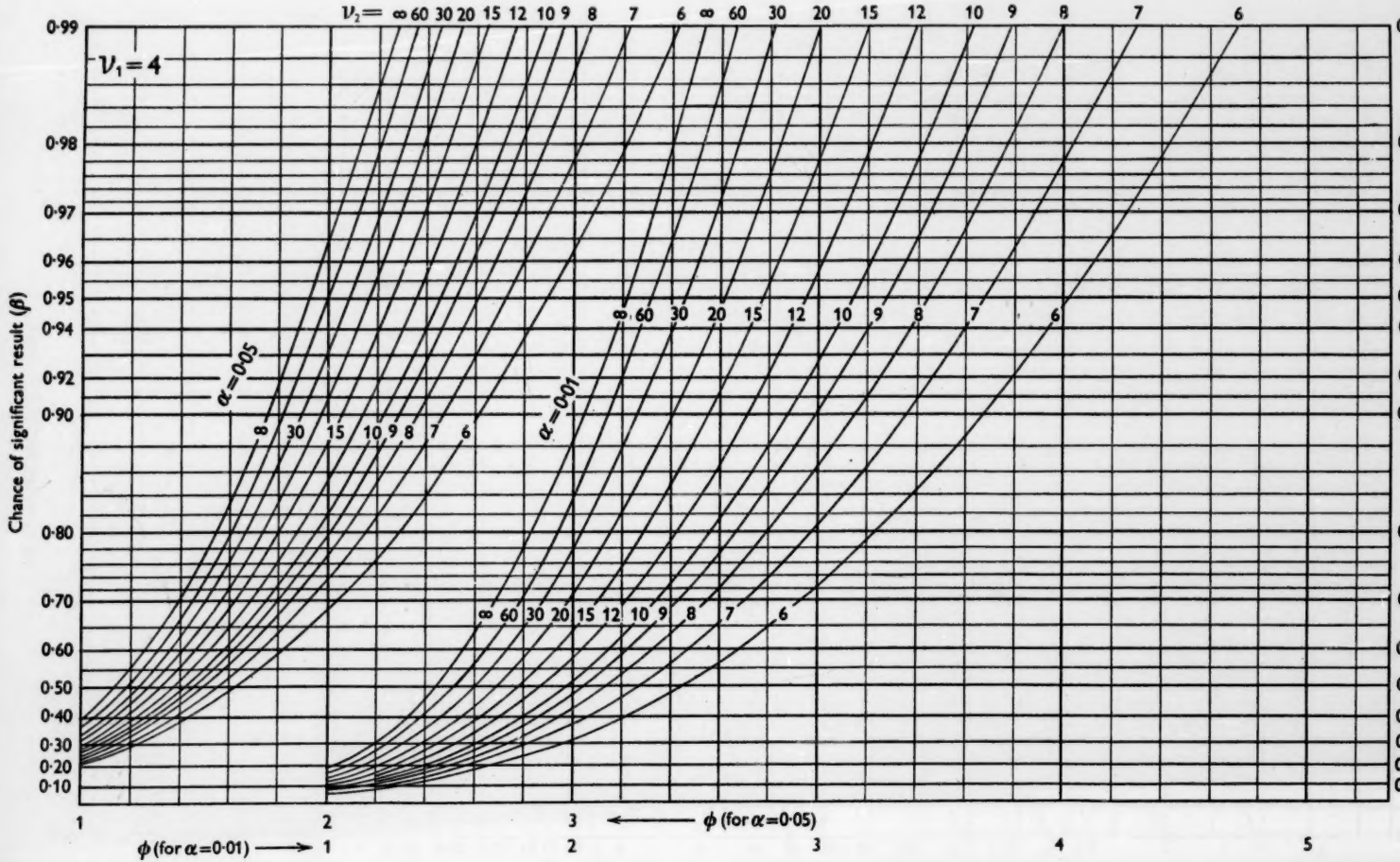
$$\phi = \sqrt{\{\lambda/(\nu_1 + 1)\}}.$$

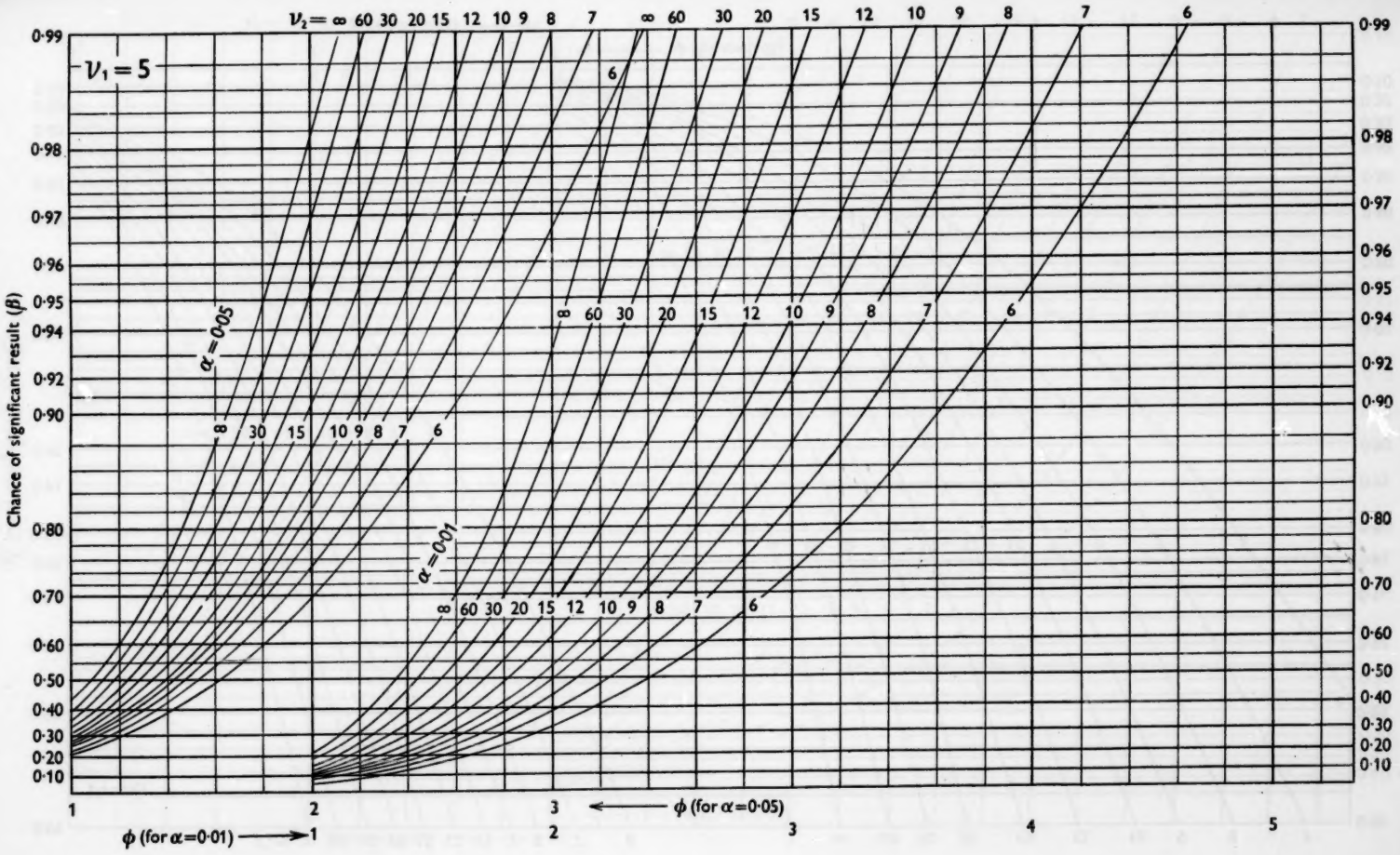
* The distinction between alternative mathematical constructions underlying the analysis of variance has been discussed by several writers, e.g. Daniels (1939), Eisenhart (1947) and Johnson (1948).

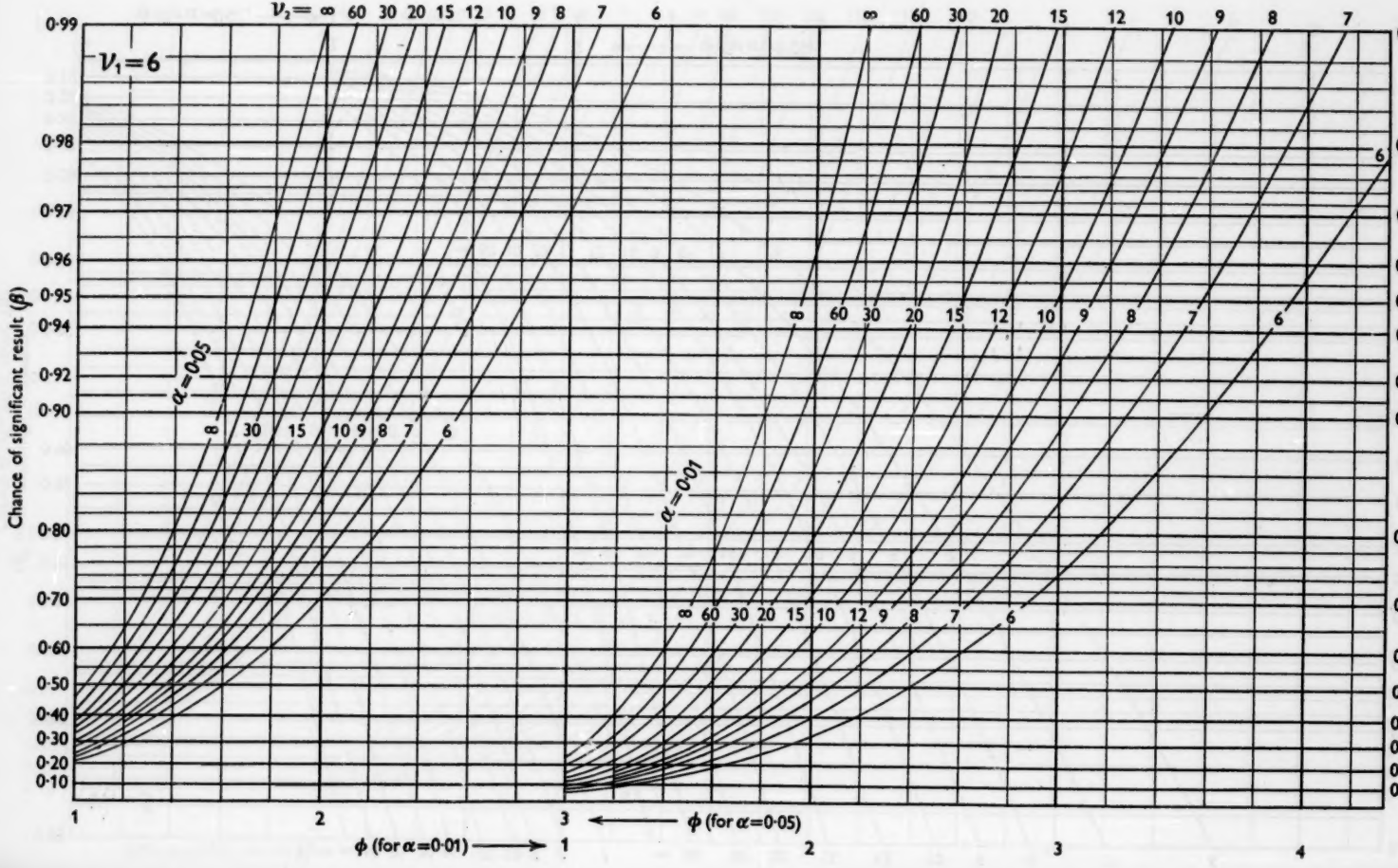

Chance of significant result (β)
$\nu_1 = 1$
$\nu_2 = \infty$ 60 30 20 15 12 10 9 8 7 6
$\alpha = 0{\cdot}01$
$\alpha = 0{\cdot}05$
ϕ (for $\alpha = 0{\cdot}01$)
ϕ (for $\alpha = 0{\cdot}05$)
0·99 0·98 0·97 0·96 0·95 0·94 0·92 0·90 0·80 0·70 0·60 0·50 0·40 0·30 0·20 0·10
1·5 2 2·5 3 3·5 4 5 6 8 10 15 30 ∞

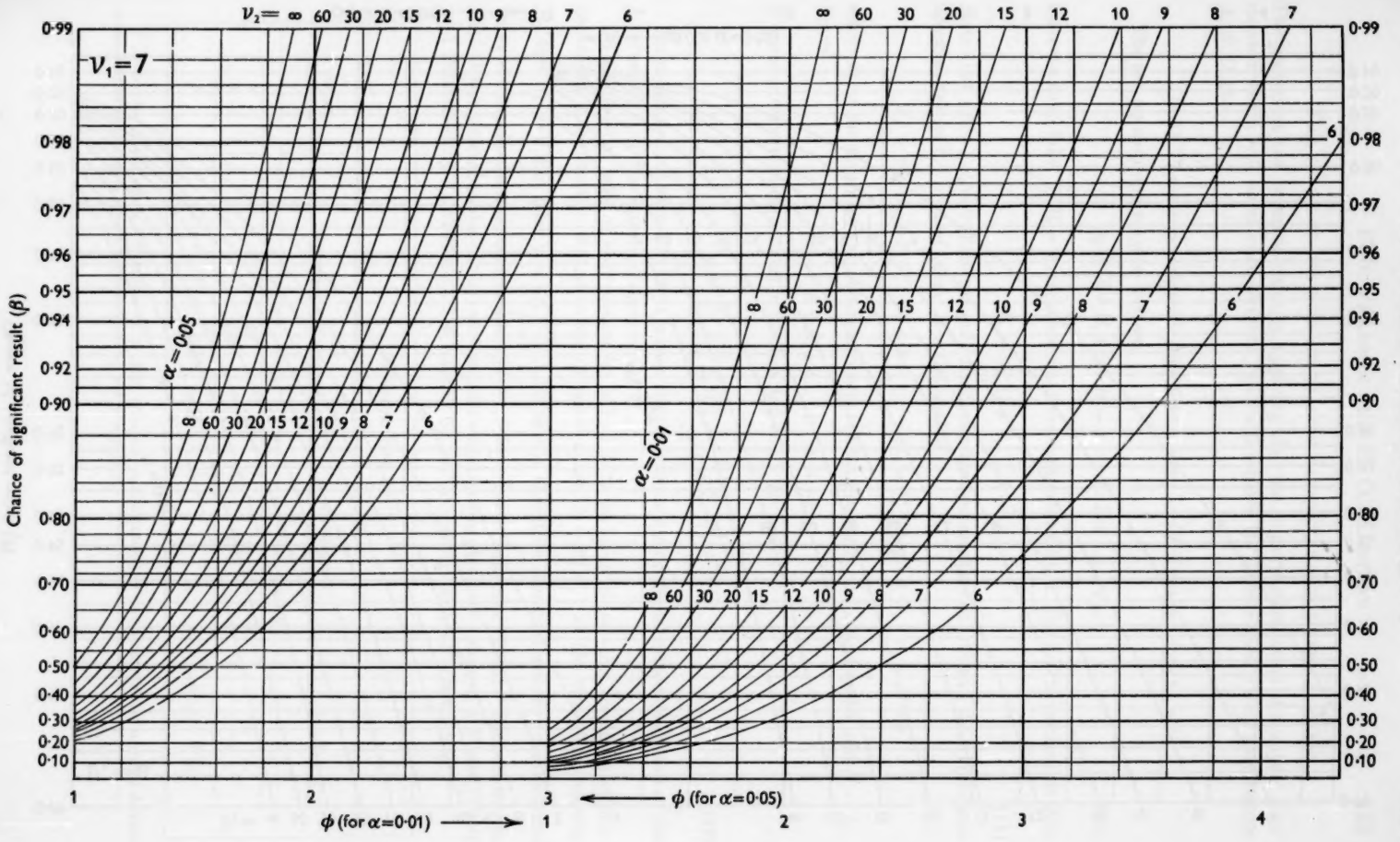

$\nu_1 = 2$
$\nu_2 =$ ∞ 60 30 20 15 12 10 9 8 7 6
$\alpha = 0{\cdot}05$
$\alpha = 0{\cdot}01$
Chance of significant result (β)
0·99
0·98
0·97
0·96
0·95
0·94
0·92
0·90
0·80
0·70
0·60
0·50
0·40
0·30
0·20
0·10
ϕ (for $\alpha = 0{\cdot}05$)
ϕ (for $\alpha = 0{\cdot}01$)
1 2 3 4 5

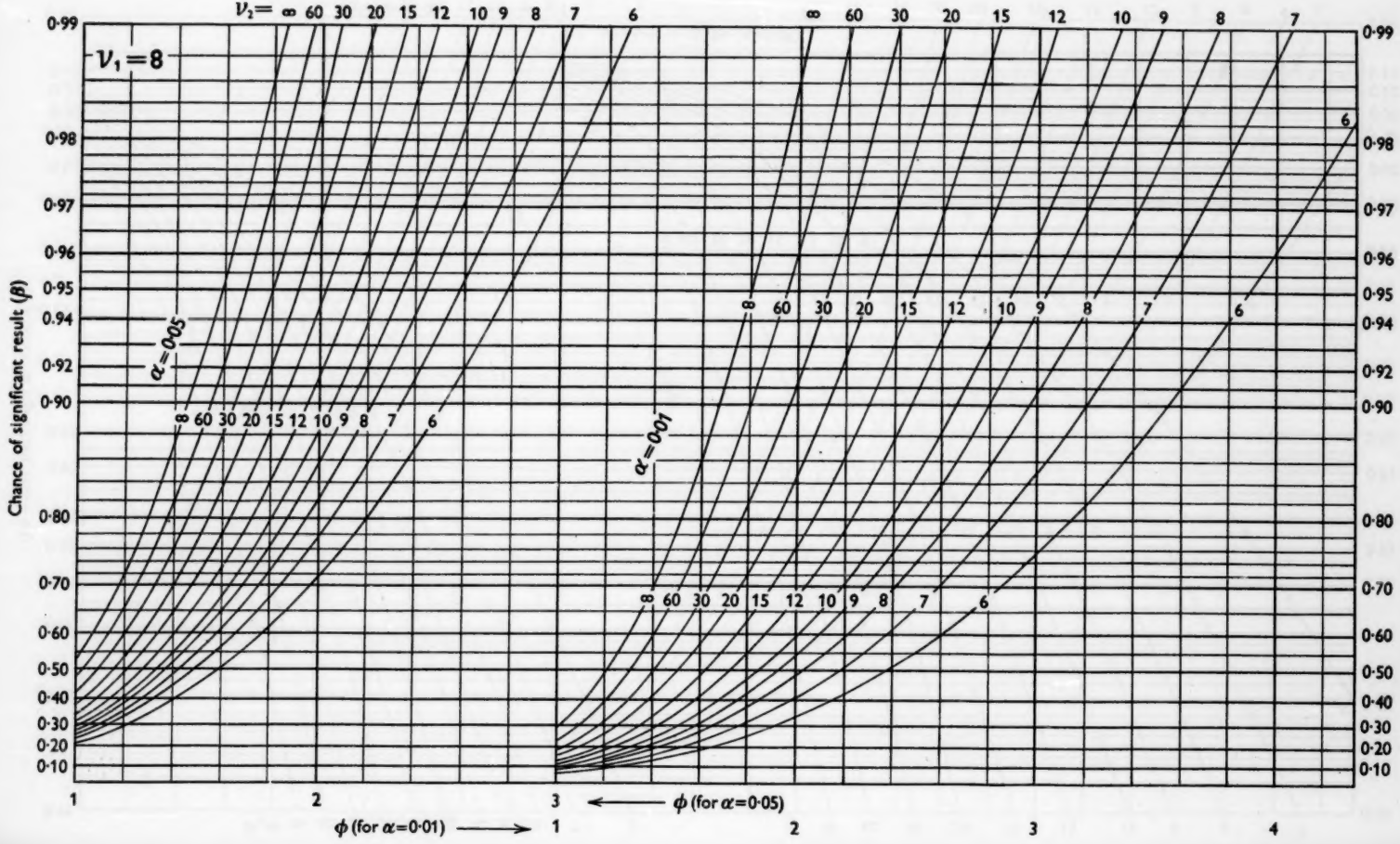

$\nu_1 = 3$
$\nu_2 =$ ∞ 60 30 20 15 12 10 9 8 7 6
$\alpha = 0{\cdot}05$
$\alpha = 0{\cdot}01$
Chance of significant result (β)
0·99
0·98
0·97
0·96
0·95
0·94
0·92
0·90
0·80
0·70
0·60
0·50
0·40
0·30
0·20
0·10
ϕ (for α=0·01) → 1 2 3
ϕ (for α=0·05) ← 2 3 4 5


$\nu_1 = 4$
$\nu_2 =$ ∞ 60 30 20 15 12 10 9 8 7 6
α=0·05
α=0·01
Chance of significant result (β)
0·99
0·98
0·97
0·96
0·95
0·94
0·92
0·90
0·80
0·70
0·60
0·50
0·40
0·30
0·20
0·10
1
2
3
4
5
φ (for α=0·01) ⟶
⟵ φ (for α=0·05)


$\nu_1 = 5$
$\nu_2 = \infty$ 60 30 20 15 12 10 9 8 7
$\alpha = 0{\cdot}05$
$\alpha = 0{\cdot}01$
Chance of significant result (β)
0·99
0·98
0·97
0·96
0·95
0·94
0·92
0·90
0·80
0·70
0·60
0·50
0·40
0·30
0·20
0·10
1
2
3
4
5
ϕ (for $\alpha = 0{\cdot}05$)
ϕ (for $\alpha = 0{\cdot}01$)


$\nu_1 = 6$
$\nu_2 = \infty$ 60 30 20 15 12 10 9 8 7 6
$\alpha = 0{\cdot}05$
$\alpha = 0{\cdot}01$
Chance of significant result (β)
0·99
0·98
0·97
0·96
0·95
0·94
0·92
0·90
0·80
0·70
0·60
0·50
0·40
0·30
0·20
0·10
1
2
3
4
ϕ (for α=0·01) ⟶
⟵ ϕ (for α=0·05)


$\nu_1 = 7$
$\nu_2 =$ ∞ 60 30 20 15 12 10 9 8 7 6
Chance of significant result (β)
0·99
0·98
0·97
0·96
0·95
0·94
0·92
0·90
0·80
0·70
0·60
0·50
0·40
0·30
0·20
0·10
$\alpha = 0{\cdot}05$
$\alpha = 0{\cdot}01$
1
2
3
4
ϕ (for $\alpha = 0{\cdot}01$) ⟶
⟵ ϕ (for $\alpha = 0{\cdot}05$)


$\nu_1 = 8$
$\nu_2 =$ ∞ 60 30 20 15 12 10 9 8 7 6
$\alpha = 0{\cdot}05$
$\alpha = 0{\cdot}01$
Chance of significant result (β)
0·99
0·98
0·97
0·96
0·95
0·94
0·92
0·90
0·80
0·70
0·60
0·50
0·40
0·30
0·20
0·10
1
2
3
4
ϕ (for $\alpha = 0{\cdot}01$) ⟶
⟵ ϕ (for $\alpha = 0{\cdot}05$)

4. THE VALUE OF THE NON-CENTRALITY PARAMETER ϕ FOR SPECIAL CASES

The procedure can be made clear by applying the rule to particular forms of analysis. For reasons which will be at once obvious, we shall replace the generalized θ's by parameters A, B, C, etc.

(4·1) *The one-way classification into k groups with n observations in each*

The set-up assumed may be written

$$x_{il} = A + B_i + z_{il} \quad (i = 1, \ldots, k;\ l = 1, \ldots, n). \tag{10}$$

Here we shall take $\sum_{i=1}^{k} B_i = 0$, representing the average level by A. The null hypothesis is that

$$B_i = 0 \quad (i = 1, \ldots, k-1). \tag{11}$$

Note that since $\Sigma B_i = 0$ there are only k independent parameters in (10) and the hypothesis specifies values for $k-1$ of them. For the analysis

$$S_1 = n\sum_i (\bar{x}_{i.} - \bar{x}_{..})^2, \quad \nu_1 = k-1; \qquad S_2 = \sum_i \sum_l (x_{il} - \bar{x}_{i.})^2, \quad \nu_2 = N-k,$$

where $\bar{x}_{i.}$ is the mean for the observations in the ith group, $\bar{x}_{..}$ the grand mean and $N = kn$. Since

$$\bar{x}_{i.} = A + B_i + \bar{z}_{i.}, \quad \bar{x}_{..} = A + \bar{z}_{..},$$

it is seen from rules (b) and (c) in § 3 that $\lambda\sigma_z^2 = n\sum_i B_i^2$, and

$$\phi = \frac{1}{\sigma_z}\sqrt{\left\{\frac{n}{k}\sum_i B_i^2\right\}} = \sqrt{\left\{\frac{1}{k}\Sigma B_i^2\right\}} \div \frac{\sigma_z}{\sqrt{n}}. \tag{12}$$

From the second form of (12) it is seen that ϕ is the ratio of the standard deviation of the k values of B_i to the standard error of the $\bar{z}_{i.}$.

In the special case of two groups ($\nu_1 = 1$, $k = 2$) we have $B_1 = -B_2$, so that the non-centrality ϕ is directly related to the distance apart, $\Delta = 2|B_1|$, of the two group means, B_1 and B_2. We find that $\phi = \frac{1}{2}\Delta\Big/\frac{\sigma_z}{\sqrt{n}}$. The test is therefore identical with the symmetrical form of the non-central t-test as described by Neyman & Tokarska (1936).* Their notation, that of the general tables of the non-central t-distribution given by Johnson & Welch (1940) and that used by Lord (1950), are related by the following table of equivalents:

Notation used by			
Present authors	Neyman & Tokarska	Johnson & Welch	Lord
α	2α	—	α
ϕ	$2^{-\frac{1}{2}}\rho$	$2^{-\frac{1}{2}}\delta$	$2^{-\frac{1}{2}}\rho$
ν_2	n	f	ν
β	$2 - P_{II}(\alpha, \rho) - P_{II}(\alpha, -\rho)$	—	$\beta'(\rho) + \beta''(\rho)$

* Note that Neyman & Tokarska's $\rho = \Delta/\sigma$, where σ is the standard deviation of the normal deviate x used in estimating Δ. Thus for $x = \bar{x}_{1.} - \bar{x}_{2.}$, $\sigma = \sigma_z(2/n)^{\frac{1}{2}}$.

The α in the last line corresponds to that of Neyman & Tokarska's notation. Since for all points plotted on our chart for $\nu_1 = 1$, $P_{II}(\alpha, -\rho) = 1$ to the plotting accuracy possible, this chart can be used also for the asymmetrical non-central t-test, using 2·5 and 0·5 % as significance levels.

If the number of observations in each group are unequal, then $\lambda\sigma_z^2 = \sum_t n_t B_t^2$. However, since the use of the power function is in connexion with advanced planning of a controlled experiment, the n_t will generally be taken equal.

(4·2) *The double classification with one observation in each cell*

Set-up:

$$x_{st} = A + B_s + C_t + z_{st} \quad (s = 1, \ldots, h;\ t = 1, \ldots, k). \tag{13}$$

Again we shall take $\sum_s B_s = 0 = \sum_t C_t$. If we are testing whether a B-effect exists, the null hypothesis is that

$$B_s = 0 \quad (s = 1, \ldots, h-1).$$

For the analysis

$$S_1 = k\sum_s (\bar{x}_{s.} - \bar{x}_{..})^2, \quad \nu_1 = h-1; \qquad S_2 = \sum_s \sum_t (x_{st} - \bar{x}_{s.} - \bar{x}_{.t} + \bar{x}_{..})^2, \quad \nu_2 = (h-1)(k-1).$$

We find at once that $\bar{x}_{s.} - \bar{x}_{..} = B_s + \bar{z}_{s.} - \bar{z}_{..}$, so that $\lambda\sigma_z^2 = k\sum_s B_s^2$ and

$$\phi = \frac{1}{\sigma_z}\sqrt{\left\{\frac{k}{h}\sum_s B_s^2\right\}} = \sqrt{\left\{\frac{1}{h}\sum_s B_s^2\right\}} \div \frac{\sigma_z}{\sqrt{k}}. \tag{14}$$

(4·3) *The double classification with n observations in each cell*

In this case it is possible to examine for the presence of interaction and the set-up may be written*

$$x_{stl} = A + B_s + C_t + D_{st} + z_{stl} \quad (s = 1, \ldots, h;\ t = 1, \ldots, k;\ l = 1, \ldots, n). \tag{15}$$

Again we shall take $\sum_s B_s = \sum_t C_t = 0$ and also

$$\sum_s D_{st} = 0 \quad (t = 1, \ldots, k) \qquad \text{and} \qquad \sum_t D_{st} = 0 \quad (s = 1, \ldots, h).$$

If we test for the presence of a main effect, say whether $B_s = 0\,(s = 1, 2, \ldots, h-1)$, we find that

$$\phi = \frac{1}{\sigma_z}\sqrt{\left\{\frac{kn}{h}\sum_s B_s^2\right\}} = \sqrt{\left\{\frac{1}{h}\sum_s B_s^2\right\}} \div \frac{\sigma_z}{\sqrt{(kn)}}. \tag{16}$$

If we test for the presence of interaction, the null hypothesis is that $D_{st} = 0$ for all combinations of s and t where, in view of the relations above, only $(h-1)(k-1)$ are independent. For the analysis,

$$S_1 = n\sum_s \sum_t (\bar{x}_{st.} - \bar{x}_{s..} - \bar{x}_{.t.} + \bar{x}_{...})^2, \quad \nu_1 = (h-1)(k-1);$$

$$S_2 = \sum_s \sum_t \sum_l (x_{stl} - \bar{x}_{st.})^2, \qquad \nu_2 = hk(n-1).$$

* We are here concerned with systematic main effects B_s, C_t and interactions D_{st}. Often it is more appropriate to regard one of the main effects (e.g. 'Blocks') and the interaction as random variables, and in this case the appropriate test procedure is different.

It is found that $\lambda\sigma_z^2 = n\sum_s\sum_t D_{st}^2$, so that

$$\phi = \frac{1}{\sigma_z}\sqrt{\left\{\frac{n}{(h-1)(k-1)+1}\sum_s\sum_t D_{st}^2\right\}}. \tag{17}$$

Since there are now hk different D_{st}, ϕ has not now the simple interpretation of a ratio of two standard deviations which it had in previous cases.

(4·4) *The Latin square*

Here, if we have a $k \times k$ square,

$$x_{rst} = A + B_r + C_s + D_t + z_{rst} \quad (r = 1, \ldots, k;\ s = 1, \ldots, k;\ t = 1, \ldots, k).$$

Let B_r and C_s represent the row and column terms and D_t the possible treatment effects. We assume

$$\sum_r B_r = \sum_s C_s = \sum_t D_t = 0$$

and the null hypothesis is that

$$D_t = 0 \quad (t = 1, 2, \ldots, k-1).$$

For the analysis

$$S_1 = k\sum_t (\bar{x}_{..t} - \bar{x}_{...})^2, \quad \nu_1 = k-1;$$

$$S_2 = \sum_r\sum_s (x_{rst} - \bar{x}_{r..} - \bar{x}_{.s.} - \bar{x}_{..t} + 2\bar{x}_{...})^2, \quad \nu_2 = (k-1)(k-2).$$

We find that $\lambda\sigma_z^2 = k\sum_t D_t^2$ and

$$\phi = \frac{1}{\sigma_z}\sqrt{\sum_t D_t^2} = \sqrt{\left\{\frac{1}{k}\sum D_t^2\right\}} \div \frac{\sigma_z}{\sqrt{k}}. \tag{18}$$

5. THE INTERPRETATION OF THE POWER FUNCTION

It will be useful at this point to consider two examples:

Example 1. *Effect of machine variation on the standard deviation of the manufactured bulk.* In a machine shop six machines contribute equal shares to the total output of mass-produced parts for which a dimension, x, is subject to tolerances. It is suspected that the average dimension of the parts produced by the six machines may differ from machine to machine, thereby increasing the standard deviation of x for the bulk. It is further feared that if such an increase should exceed 20 % of the standard deviation of x for parts from a single machine previously employed, difficulties would arise in meeting the tolerances.

In order to have advance warning of machine differences of such a magnitude, it is planned to measure nine parts for each of the six machines and test for machine differences in an analysis of variance, using the 1 % level of significance. What is the chance of detecting machine differences of such a magnitude by this test?

It would seem reasonable to adopt the set-up (4·1). Let $A + B_t$ denote the average dimension of parts produced by the tth machine, x_{tl} the dimension of the lth part measured for the tth machine and σ the standard deviation of dimensions of parts from the same machine. Then it is desired to detect by the test the presence of differences among the B_t such that

$$\sqrt{\{\tfrac{1}{6}\Sigma B_t^2 + \sigma^2\}}/\sigma \geqslant 1{\cdot}2 \quad (\text{increase of } 20\ \% \text{ in } \sigma),$$

or that

$$\sqrt{\{\tfrac{1}{6}\Sigma B_t^2\}}/\sigma \geqslant \sqrt{(1{\cdot}2^2 - 1)} = \sqrt{0{\cdot}44} = 0{\cdot}66.$$

Hence $\phi = 0{\cdot}66\sqrt{9} = 1{\cdot}99$ is the critical value.

Entering the chart for $\nu_1 = 6-1 = 5$ and the section for $\alpha = 0{\cdot}01$ (1 % level of significance), we find for $\phi = 1{\cdot}99$ and $\nu_2 = 6(9-1) = 48$ the value of $\beta = 0{\cdot}86$. This means that, following the procedure described, our chance of detecting machine differences of the above magnitude is 0·86, i.e. the odds are about 6 to 1 on detection. Should these not be considered sufficient, we must make provision for more parts to be measured, until a satisfactory value of β is reached.

Whilst the planning of the present experiment was concerned with the value of the root-mean-square of the group means B_t, situations may of course arise in which the values of individual B_t are of interest.

Example 2. *The effect of personal factors introduced by test operators in certain routine tests.* Davies (1947, p. 90) has described an investigation into the possible influence of personal factors on the results of a standard procedure of testing the compressive strength of samples of Portland cement. Two classes of men are employed in the test: the 'gauger' who works up the mixture of cement and water from which the test cubes are moulded, and the assistant who operates the testing machine, who is described as a 'breaker'. Let us suppose, as in the case of the data given, that there were three assistants regularly employed on the testing and three labourers who were used for the gauging. To investigate the possible effects of personal factors among the men, we may plan a simple 3×3 experiment in which n cubes are tested for each of the nine combinations of gauger and breaker. If we were concerned with six specified men who were to be employed on this work in a routine way, it would seem reasonable to adopt the set-up of § (4·3) above. B_1, B_2 and B_3 will represent the possible gauger effects and C_1, C_2 and C_3 the breaker effects. In testing for the presence of either of these, we shall have

$$\nu_1 = 2, \quad \nu_2 = 9(n-1),$$

while from equation (16)

$$\phi_B = \sqrt{\{n\Sigma B_s^2\}}/\sigma_z,$$

and similaly for ϕ_C. If we were to decide to take $n = 4$, so that $\nu_2 = 27$, we can derive, for example, the following information from the chart for $\nu_1 = 2$:

(*a*) The probability will be at least 0·90 of establishing significance for a gauger (or breaker) effect at the 5 % level of F if

$$\phi \geqslant 2{\cdot}16 \quad \text{or if} \quad \sqrt{\{\tfrac{1}{3}\sum_s B_s^2\}} \geqslant \frac{2{\cdot}16}{\sqrt{12}}\sigma_z = 0{\cdot}62\sigma_z.$$

(*b*) To have the same probability, using the 1 % significance level, we must have

$$\phi \geqslant 2{\cdot}63 \quad \text{or} \quad \sqrt{\{\tfrac{1}{3}\sum_s B_s^2\}} \geqslant 0{\cdot}76\sigma_z.$$

(*c*) If, however, the values of the men's averages were less scattered, i.e. only such that

$$\sqrt{\{\tfrac{1}{3}\sum_s B_s^2\}} = 0{\cdot}5\sigma_z \quad \text{or} \quad \phi = \sqrt{(12)}\,0{\cdot}5 = 1{\cdot}73,$$

then there would be probabilities of only (i) 0·72 and (ii) 0·45 of establishing significance at (i) the 5 % and (ii) the 1 % level.

(*d*) Were the number of tests for each gauger-breaker combination increased from $n = 4$ to $n = 8$, so that ν_2 becomes 63, then, still using the same chart, we find that $\sqrt{\{\tfrac{1}{3}\sum_s B_s^2\}}$ need only be (i) $0{\cdot}43\sigma_z$ and (ii) $0{\cdot}51\sigma_z$ to give a probability of 0·90 of establishing significance at (i) the 5 % and (ii) the 1 % level.

This example brings out two points needing further consideration. In the first place the root-mean-square of the group-means B_i is expressed in terms of σ_z, the standard deviation of the normal residuals. It will often happen that this is roughly known in advance of the experiment. For example, it might have been known in advance from past experience that if a breaker tests a series of cement cubes made from the same mix, the standard deviation of his results would be about 500 lb./sq.in. Having this information, rough numerical figures could be assigned to the critical values for $\sqrt{\{\frac{1}{3}\Sigma B_j^2\}}$ quoted above. Thus for case (a) we have

$$\sqrt{\{\tfrac{1}{3}\Sigma B_j^2\}} = 0{\cdot}62 \times 500 = 310 \text{ lb./sq.in.}$$

The second point is this. How far does the root-mean-square of the B_i provide the kind of information we need in a preliminary survey? To generalize for the purpose of the present discussion, write the relevant factors B_s, C_t, ..., etc. as δ_j ($j = 1, \ldots, k$), where

$$\sum_j \delta_j = 0, \quad \sum_j \delta_j^2 = c^2\phi^2\sigma_z^2, \tag{19}$$

the constant c depending on the arrangement and being given for particular cases by (12), (14), (16), (17) and (18) above.* For a given value of the non-centrality parameter, whether expressed as λ or ϕ or $\Sigma\delta_j^2$, there will be an infinite set of possible combinations of the δ_j. Can we pick out some characteristics of this set to which we can attach particular meaning?

In some unpublished notes completed before he returned to China in 1938, Tang suggested the following approach. For every set of values of the δ_j satisfying (19) there will be a value which is numerically greatest. This set of greatest numerical values must be bounded; he wrote the lower and upper bounds as follows:

$$\Delta_L = \text{the least upper deviation},$$

$$\Delta_G = \text{the greatest upper deviation},$$

and showed that

$$\left.\begin{aligned} \Delta_L &= \begin{cases} c\phi\sigma_z/\sqrt{k} & k \text{ even}, \\ c\phi\sigma_z/\sqrt{(k-1)} & k \text{ odd}, \end{cases} \\ \Delta_G &= c\phi\sigma_z\sqrt{\{(k-1)/k\}}. \end{aligned}\right\} \tag{20}$$

The meaning of this result he interpreted as follows:

For given values of ϕ and σ_z (c being defined by the experimental arrangement) there must be at least one δ_j whose numerical value is as great or greater than Δ_L, but there can be no δ_j whose numerical value is greater than Δ_G.

Δ_L and Δ_G are clearly not the only characteristics of interest. We are inclined to think that the range or spread of the δ_j may be even more relevant. We will denote by $W(\text{max.})$ and $W(\text{min.})$ the upper and lower bounds of the ranges of the combination of δ's satisfying (19) for fixed ϕ and σ_z. Then it can be shown that

$$W(\text{max.}) = c\phi\sigma_z \times \sqrt{2}, \tag{21}$$

$$W(\text{min.}) = \begin{cases} c\phi\sigma_z \times 2/\sqrt{k} & k \text{ even}, \\ c\phi\sigma_z \times 2\sqrt{\{k/(k^2-1)\}} & k \text{ odd}. \end{cases} \tag{22}$$

The maximum range occurs when the extreme δ_j are $\pm c\phi\sigma_z/\sqrt{2}$ and the remaining $k-2$ zero.

* This form of relation between the δ's and ϕ is appropriate for all the arrangements considered in § 4, but it will not always be so in tests of the general linear hypothesis. For example, if there are unequal frequencies in subgroups, the weighted sum of squares of the δ's is involved.

The minimum range occurs, (i) if k is even, when $\frac{1}{2}k$ of the δ_j equal $\pm c\phi\sigma_z/\sqrt{k}$ and the remaining $\frac{1}{2}k$ of the δ_j equal $\mp c\phi\sigma_z/\sqrt{k}$; (ii) if $k = 2p+1$ is odd, when $p+1$ of the δ_j equal

$$\pm c\phi\sigma_z\{(k-1)/(k+1)\,k\}^{\frac{1}{2}},$$

and the remaining p equal $\quad \mp c\phi\sigma_z\{(k+1)/(k-1)\,k\}^{\frac{1}{2}}$.

The interpretation is as follows. For given values of ϕ and σ_z, the range of the δ_j, i.e. the spread in the expectations of the x's resulting from treatment differences, must be at least W(min.) but cannot exceed W(max.).* Alternatively, given W and σ_z, we can invert the relations (21) and (22) to obtain lower and upper bounds for ϕ and therefore for the power of the test. These points may be illustrated on the example considered on pp. 126–7 above. Here, for the differences between gaugers,

$$\sum_{s=1}^{3} B_s^2 = \frac{1}{n}\phi^2\sigma_z^2,$$

or the c of equations (19), (21) and (22) is $c = 1/\sqrt{n}$. Suppose that having regard to the general programme of routine testing and the conclusions to be based upon it, it was considered that the maximum spread of what may be termed the gauger bias should not be more than $W = 250$ lb./sq.in. Further, suppose again that our rough estimate of σ_z is 500 lb./sq.in.

(*a*) We might first ask if there is any prospect of establishing significance at the 5 % level when $W = 250$ lb./sq.in. if a 3×3 experimental arrangement were used with only $n = 4$ replications. $k = 3$ and is odd, so from (22) we see that

$$\phi(\text{max.}) = \frac{W}{2\sigma_z}\sqrt{\left\{\frac{n(k^2-1)}{k}\right\}} \tag{23}$$

$$= \frac{250}{2 \times 500}\sqrt{\frac{4 \times 8}{3}} = \sqrt{\frac{2}{3}} = 0{\cdot}816.$$

Turning to the chart for $\nu_1 = 2$, with $\nu_2 = 27$, it is clear that the value of ϕ is outside the range of both charts. Extrapolation suggests that there is a probability of only about 0·25 of establishing significance at the 5 % level. Thus four replications would be quite inadequate to detect a spread of 250 lb./sq.in.

(*b*) As indicated under (*a*) on p. 126 above, for the probability of establishing significance at the 5 % level to be 0·90 or more, we must have $\phi \geqslant 2{\cdot}16$. Taking W and σ_z as before we have from (21)

$$\phi(\text{min.}) = \frac{W}{\sigma_z}\sqrt{\frac{n}{2}} = \frac{\sqrt{n}}{2\sqrt{2}}. \tag{24}$$

Thus, to ensure that the probability of establishing significance at the 5 % level is 0·90 or more, we must make $\sqrt{n}/(2\sqrt{2}) \geqslant 2{\cdot}16$ or $n \geqslant 37{\cdot}3$. Thus 38 replications would be needed.

(*c*) Suppose now that we were content with a reasonable chance of establishing significance if the spread in gauger bias were $W = 500$ lb./sq.in. From (23) and (24) we have

$$\phi(\text{max.}) = 0{\cdot}816\sqrt{n}, \quad \phi(\text{min.}) = 0{\cdot}707\sqrt{n}.$$

* Whilst for small k the upper and lower bounds W(max.) and W(min.) are not too far apart and may be used as indicated in the example below, their increasing divergence for the larger k may make the interpretation illustrated less useful. In such cases it may be possible to interpret the practical significance of a value of ϕ directly (as in Example 1) or to make some additional assumptions on the distribution of the δ_j.

Taking $n = 4$ and examining the chart for $\nu_1 = 2$ again, we find that the probability of establishing significance at the 5 % level lies between 0·66 and 0·52. The former will result if, say, $B_1 = B_2 = 167$ lb./sq.in. and $B_3 = -333$ lb./sq.in.; and the latter if $B_1 = -B_3 = 250$ lb./sq.in., $B_2 = 0$.

(*d*) For the probability to be at least 0·90 the number of replications must satisfy the inequality $n \geqslant (2{\cdot}16/0{\cdot}707)^2 = 9{\cdot}3$ or 10 replications would suffice.

6. COMPUTATION FOR THE CHARTS

The basic formula for the computation of the present charts was an approximation to the non-central F-distribution, defined by (5). Patnaik (1949) showed that, to a fair degree of accuracy, the distribution of non-central F' based on ν_1 and ν_2 degrees of freedom could be represented by that of $cF(\nu, \nu_2)$, where $F(\nu, \nu_2)$ is the ordinary central F-ratio in which the degrees of freedom of the denominator of F' have been retained, whilst the degrees of freedom (ν) of the numerator are given by

$$\nu = (\nu_1 + \lambda)^2/(\nu_1 + 2\lambda),$$

and the scale factor c is given by

$$c = (\nu_1 + \lambda)/\nu_1.$$

Using this approximation and the familiar formula for the F-integral in terms of the Incomplete B-function, viz.

$$\beta = I_x(\tfrac{1}{2}\nu_2, \tfrac{1}{2}\nu), \quad x = \nu_2/(\nu_2 + \nu F),$$

values of $\beta(\phi, \nu_1, \nu_2)$ were computed from K. Pearson's *Tables of the Incomplete Beta-function* for a grid of values for ϕ, ν_1, ν_2 which included the rather wide grid in Tang's table. At these latter grid points a comparison with Tang's results provided a check on Patnaik's approximation. It was found, however, that in order to determine the correction required for the remaining values of the grid computed from Patnaik's approximation, it was necessary to compute further exact values, particularly for large β, where our logarithmic scale necessitates high accuracy. For this purpose Tang's recurrence formula was used for small ν_2 ($\nu_2 = 6$–8), whilst for large ν_2 a new formula for non-central χ^2 provided by Mr F. W. J. Olver, of the National Physical Laboratory, proved helpful. With the help of this formula small tables of non-central χ^2 were computed and 'Studentized' according to Hartley's formulae (1944), in which the differentials of non-central χ^2 were replaced by their finite-difference expansions. This method provided check answers for $\nu_2 = 20$, 30, 60 and ∞.

With the help of these further exact values a table of corrections,

$$\beta(\text{exact}) - \beta(\text{Patnaik's approximation}),$$

could be prepared, which permitted rough interpolation and thus enabled the remaining values of the grid to be corrected to an accuracy sufficient for the purpose of plotting the present charts. Further plotting points were provided by the non-central F-table recently computed by Lehmer (1944) for $\beta = 0{\cdot}7$ and 0·8. All values for $\nu_1 = 1$ were computed directly from Johnson & Welch's (1940) table of non-central-t, and many of those for $\nu_2 = \infty$ from a table of non-central χ^2 recently computed by Fix (1949).

For the organization of the computational work we are again indebted to Mr T. Vickers of the Mathematics Division of the National Physical Laboratory, as well as to Miss J. H. Thompson of the Department of Statistics, University College.

The suggestion that Tang's tables might be put into more convenient form for practical use was originally made to us by Dr G. P. Sillitto of Imperial Chemical Industries Ltd., and we are indebted to that firm for a substantial grant towards the cost of computation.

REFERENCES

DANIELS, H. E. (1939). *J.R. Statist. Soc. Suppl.*, **6**, 186.
DAVIES, O. L. (1947). *Statistical Methods in Research and Production.* Edinburgh: Oliver and Boyd.
EISENHART, C. (1947). *Biometrics*, **3**, 1.
FISHER, R. A. (1928). *Proc. Roy. Soc.* A, **121**, 654.
FIX, EVELYN (1949). *Univ. Calif. Publ. Statist.* **1**, 15.
HARTLEY, H. O. (1944). *Biometrika*, **33**, 173.
JOHNSON, N. L. (1948). *Biometrika*, **35**, 80.
JOHNSON, N. L. & WELCH, B. L. (1940). *Biometrika*, **31**, 362.
KOLODZIEJCZYK, S. (1935). *Biometrika*, **27**, 161.
LEHMER, EMMA (1944). *Ann. Math. Statist.* **15**, 388.
LORD, E. (1950). *Biometrika*, **37**, 64.
NEYMAN, J. & TOKARSKA, B. (1936). *J. Amer. Statist. Ass.* **31**, 318.
PATNAIK, P. B. (1949). *Biometrika*, **36**, 202.
TANG, P. C. (1938). *Statist. Res. Mem.* **2**, 126.

SOME QUESTIONS OF DISTRIBUTION IN THE THEORY OF RANK CORRELATION

By S. T. DAVID, M. G. KENDALL and A. STUART

Division of Research Techniques, London School of Economics

Summary

1. This paper deals with the following matters:

(*a*) We give the actual distribution of Spearman's ρ_s in the null case for rankings of 9 and 10.

(*b*) We give the moments and cumulants of Spearman's ρ_s in the null case up to those of the eighth order.

(*c*) We use these cumulants and the cumulants of τ recently given by Silverstone (1950) to develop an expansion of the frequency functions of ρ_s and τ in the null case.

(*d*) We examine the product-moment correlation between r_s and t in samples from a bivariate normal population.

2. A valuable review of recent work on ranking methods has been given by Moran (1950) as part of a symposium to which Whitfield (1950) and Daniels (1950) contributed. These authors called attention to certain undesirable features of the notation of the subject, but as they all used different notations themselves and some of the contributors to the discussion used different notations again, we shall adhere to the notation of a previous paper by one of us (Kendall, 1949), that is to say: Spearman's ρ will be denoted by ρ_s (or r_s if sample values are in question); the alternative coefficient will be denoted by τ (or t in sample values); and the correlation parameter in a bivariate normal population will be denoted by ρ.

The frequency function of ρ_s

3. The frequency distribution of ρ_s in the null case (i.e. in the population of rankings in which each ranking occurs equally frequently) was given by Kendall and others in 1939 for n, the rank number, from 2 to 8 inclusive. Olds (1938) had independently given the distributions for $n = 2$ to 7 inclusive. The method of obtaining the distributions used by both authors is essentially the same and consists of expanding a permanent. The work of explicit expansion rapidly increases as n becomes larger, and we have not found any methods of alleviating it other than those given by Kendall *et al.* in their 1939 paper. The permanent has now been expanded for $n = 9$ and $n = 10$, which is about as far as a computer's patience can be expected to extend, and the resulting distributions are given in Table 1. The figures were checked by calculating the second and fourth moments, which are known as polynomials in n. Table 2 gives the corresponding single-tailed probabilities.

4. The frequency polygons of the distributions exhibit the characteristic local irregularities commented upon by Kendall and others and, so far as they go, confirm their conjecture that as n becomes larger (and the distribution tends to normality) the tails of the distribution smooth out first.

We are indebted to Miss Joan Ayling, of the National Institute for Social and Economic Research, who helped in carrying out these calculations with her customary patience and accuracy.

Table 1. *Distribution of $S(d^2)$ (for ρ_s) in the null case for $n = 9$ and $n = 10$*

Values of $S(d^2)$	Value of frequency		Values of $S(d^2)$	Frequency		Values of $S(d^2)$	Frequency	
	$n = 9$	$n = 10$		$n = 9$	$n = 10$		$n = 9$	$n = 10$
0	1	1	60	2878	9892	120	6688*	38456
2	8	9	62	2928	10678	122	—	38926
4	21	28	64	3397	11647	124	—	39984
6	34	51	66	3138	12141	126	—	39386
8	72	107	68	3647	13026	128	—	42068
10	102	177	70	3568	13918	130	—	42848
12	130	234	72	3921	14519	132	—	42424
14	190	360	74	3866	15611	134	—	42925
16	260	498	76	4311	16278	136	—	45044
18	284	619	78	4050	16780	138	—	44584
20	398	819	80	4852	18686	140	—	46040
22	454	1040	82	4492	19280	142	—	45496
24	555	1252	84	4816	19586	144	—	46890
26	616	1528	86	4784	20795	146	—	47036
28	756	1824	88	5505	22385	148	—	47341
30	744	2010	90	4954	22896	150	—	47646
32	1022	2533	92	5638	23948	152	—	48887
34	1042	2837	94	5304	24970	154	—	48840
36	1159	3180	96	5890	26012	156	—	48044
38	1282	3676	98	5486	27096	158	—	48540
40	1555	4305	100	6188	28467	160	—	50066
42	1392	4493	102	5502	28427	162	—	48970
44	1719	5130	104	6436	30540	164	—	49062†
46	1758	5672	106	5822	31272	166	—	49062†
48	2009	6156	108	6233	31774			
50	2032	6909	110	6024	33264			
52	2282	7424	112	6697	34748			
54	2214	7830	114	5720	34499			
56	2676	8773	116	6672	36299			
58	2590	9392	118	6020	36596			
Totals (of whole distribution)							362880	3628800

* Modal value. Values of higher S are obtainable by symmetry.
† Modal values. Values of higher S are obtainable by symmetry.

Moments and cumulants of ρ_s and τ

5. Moran (1950) and independently Silverstone (1950) have given an expression for the cumulants of τ in the null case. Silverstone's form for S, which is related to τ by the equation

$$\tau = \frac{S}{\frac{1}{2}n(n-1)},$$

is as follows:

$$(-1)^{k+1}\kappa_{2k} = \frac{2^{2k-1}}{k} B'_{2k} \sum_{j=1}^{n} (j^{2k} - 1) \tag{1}$$

$$= \frac{2^{2k-1}}{k} B'_{2k} \left\{ \frac{B_{2k+1}(n)}{2k+1} + n^{2k} - n \right\}, \tag{2}$$

Table 2. *Probability that $S(d^2)$ (for ρ_s) will be attained or exceeded for $n = 9$ and $n = 10$*

Values of $S(d^2)$	Probability		Values of $S(d)^2$	Probability		Values of $S(d^2)$	Probability	
	$n = 9$	$n = 10$		$n = 9$	$n = 10$		$n = 9$	$n = 10$
0	1·00000	1·00000	60	0·9191	0·9755	120	0·509	0·786
2	$0{\cdot}9^5724$	$0{\cdot}9^6724$	62	0·9112	0·9728	122	0·491*	0·776
4	$0{\cdot}9^4752$	$0{\cdot}9^5724$	64	0·9031	0·9698	124	—	0·765
6	$0{\cdot}9^4173$	$0{\cdot}9^4895$	66	0·894	0·9666	126	—	0·754
8	$0{\cdot}9^3824$	$0{\cdot}9^4755$	68	0·885	0·9633	128	—	0·743
10	$0{\cdot}9^3625$	$0{\cdot}9^4460$	70	0·875	0·9597	130	—	0·732
12	$0{\cdot}9^3344$	$0{\cdot}9^3897$	72	0·865	0·9559	132	—	0·720
14	$0{\cdot}9^3899$	$0{\cdot}9^3833$	74	0·854	0·9519	134	—	0·708
16	$0{\cdot}9^3846$	$0{\cdot}9^3734$	76	0·844	0·9476	136	—	0·696
18	$0{\cdot}9^2775$	$0{\cdot}9^3596$	78	0·832	0·9431	138	—	0·684
20	$0{\cdot}9^2696$	$0{\cdot}9^3426$	80	0·821	0·9384	140	—	0·672
22	$0{\cdot}9^2587$	$0{\cdot}9^3200$	82	0·807	0·9333	142	—	0·659
24	$0{\cdot}9^2462$	$0{\cdot}9^2891$	84	0·795	0·9280	144	—	0·646
26	$0{\cdot}9^2309$	$0{\cdot}9^2857$	86	0·782	0·9226	146	—	0·633
28	$0{\cdot}9^2139$	$0{\cdot}9^2815$	88	0·769	0·9169	148	—	0·621
30	0·9893	$0{\cdot}9^2764$	90	0·753	0·9107	150	—	0·607
32	0·9873	$0{\cdot}9^2709$	92	0·740	0·9044	152	—	0·594
34	0·9844	$0{\cdot}9^2639$	94	0·724	0·898	154	—	0·581
36	0·9816	$0{\cdot}9^2561$	96	0·710	0·891	156	—	0·567
38	0·9784	$0{\cdot}9^2473$	98	0·693	0·884	158	—	0·554
40	0·9748	$0{\cdot}9^2372$	100	0·678	0·876	160	—	0·541
42	0·9706	$0{\cdot}9^2254$	102	0·661	0·868	162	—	0·527
44	0·9667	$0{\cdot}9^2130$	104	0·646	0·861	164	—	0·514
46	0·9620	0·9899	106	0·628	0·852	166	—	0·500†
48	0·9571	0·9883	108	0·612	0·844			
50	0·9516	0·9866	110	0·595	0·835			
52	0·9460	0·9847	112	0·578	0·826			
54	0·9397	0·9827	114	0·560	0·816			
56	0·9336	0·9805	116	0·544	0·807			
58	0·9262	0·9781	118	0·526	0·797			

* For greater values of $S(d^2)$ the probability is the complement of the probability corresponding to $242 - S(d^2)$, e.g. for $S(d^2) = 200$ is 0·0294.

† For greater values of $S(d^2)$ the probability is the complement of the probability corresponding to $332 - S(d^2)$, e.g. for $S(d^2) = 200$ is 0·280.

where $B_j(n)$ is the jth Bernoulli polynomial in n and B'_j is the jth Bernoulli number. The actual values up to and including the cumulants of order eight (odd-order cumulants being zero) are

$$\kappa_2 = \tfrac{1}{15}n(n-1)(2n+5), \tag{3}$$

$$\kappa_4 = -\tfrac{1}{225}n(6n^4+15n^3+10n^2-31), \tag{4}$$

$$\kappa_6 = \tfrac{8}{1323}n(6n^6+21n^5+21n^4-7n^2-41), \tag{5}$$

$$\kappa_8 = -\tfrac{8}{675}n(10n^8+45n^7+60n^6-42n^4+20n^2-93). \tag{6}$$

6. The second moment of ρ_s in the null case was obtained by 'Student' about forty years ago, and the fourth moment was obtained by Hotelling & Pabst (1936). There does not appear to be any simple generating function for either moments or cumulants, and we have obtained

the moments by ascertaining expected values in a manner which is exemplified by the following derivation of μ_4.

If x and y denote two rankings of n measured from their mean value $\frac{1}{2}(n+1)$, Spearman's ρ_s may be defined as

$$\rho_s = \frac{\Sigma xy}{\frac{1}{12}n(n^2-1)}, \tag{7}$$

where x and y vary from $-\frac{1}{2}(n-1)$ to $\frac{1}{2}(n-1)$. The determination of the moments of ρ_s about the mean is then equivalent to finding the expectations of powers of Σxy. We have

$$E(\Sigma xy)^4 = E(\Sigma x^4y^4 + 4\Sigma x^3xy^3y + 3\Sigma x^2x^2y^2y^2 + 6\Sigma x^2xxy^2yy + \Sigma xxxxyyyy),$$

where we write Σx^2xx, for instance, to represent $\Sigma x_i^2 x_j x_k$ summed over all unequal values of i, j, k from $-\frac{1}{2}(n-1)$ to $\frac{1}{2}(n-1)$. In terms of augmented symmetric functions (David & Kendall, 1949) we may write, since x and y vary independently in the null case

$$\begin{aligned} &E\Sigma x^{\alpha_1}y^{\alpha_1}x^{\alpha_1}y^{\alpha_1}\dots x^{\alpha_r}y^{\alpha_r} \\ &\quad = n^{(\Sigma p_i)}E(x^{\alpha_1}y^{\alpha_1}\dots x^{\alpha_r}y^{\alpha_r}) \\ &\quad = [\alpha_1^{p_1}\alpha_2^{p_2}\dots\alpha_r^{p_r}]^2/n^{(\Sigma p_i)} \end{aligned} \tag{8}$$

where there are p_i superscripts α_i, and hence

$$E(\Sigma xy)^4 = \frac{[4]^2}{n} + \frac{4[31]^2}{n^{(2)}} + \frac{3[2^2]^2}{n^{(2)}} + \frac{6[21^2]^2}{n^{(3)}} + \frac{[1^4]}{n^{(4)}}. \tag{9}$$

We now express the symmetric functions in terms of monomial symmetrics, using for the purpose the tables of David & Kendall (1949), e.g.

$$[4] = (4),$$
$$-[2^2] = (4)-(2)^2, \text{ etc.}$$

We also have by summing the appropriate sums of powers

$$(4) = \frac{3n^2-7}{20}\frac{n(n^2-1)}{12},$$
$$(2) = \frac{n(n^2-1)}{12}, \text{ etc.}$$

On substitution in (9) we then find, after some purely algebraical reduction,

$$E(\Sigma xy)^4 = \frac{(n-1)n^2(n+1)^2}{12^2.100}(25n^4-13n^3-73n^2+37n+72),$$

whence

$$E(\rho^4) = \frac{3(25n^3-38n^2-35n+72)}{25n(n-1)^3(n+1)}.$$

The algebra rapidly becomes more complicated as the order of the moment increases. The calculation of the eighth moment involved numbers of thirteen digits.

7. Proceeding in this way we find for the even-order moments of ρ_s

$$\mu_2 = \frac{1}{n-1}, \tag{10}$$

$$\mu_4 = \frac{3(25n^3-38n^2-35n+72)}{25n(n+1)(n-1)^3}, \tag{11}$$

$$\begin{aligned} \mu_6 = \frac{3}{245n^3(n-1)^5(n+1)^3}\{&1{,}225n^8-4{,}361n^7-178n^6+23{,}818n^5 \\ &-22{,}783n^4-50{,}081n^3+54{,}280n^2+44{,}160n-28{,}800\}, \end{aligned} \tag{12}$$

$$\mu_8 = \frac{3}{875n^5(n+1)^5(n-1)^7}\{30,625n^{13} - 218,050n^{12} + 451,718n^{11} + 1,090,534n^{10} - 6,275,976n^9 + 2,142,858n^8 + 30,402,746n^7 - 27,330,110n^6 - 79,689,881n^5 + 71,871,632n^4 + 110,888,256n^3 - 74,721,024n^2 - 51,867,648n + 40,642,560\}. \quad (13)$$

From these results we have for the cumulants of ρ_s

$$\kappa_2 = \frac{1}{n-1}, \quad (14)$$

$$\kappa_4 = -\frac{6(19n^2+5n-36)}{25n(n+1)(n-1)^3}, \quad (15)$$

$$\kappa_6 = \frac{48(583n^6+723n^5-2,603n^4-2,637n^3+4,054n^2+2,760n-1,800)}{245n^3(n-1)^5(n+1)^3}, \quad (16)$$

$$\kappa_8 = \frac{144}{875n^5(n+1)^5(n-1)^7}\{-41,939n^{10} - 83,709n^9 + 304,254n^8 + 578,442n^7 - 1,012,323n^6 - 1,690,125n^5 + 1,800,776n^4 + 2,358,048n^3 - 1,616,688n^2 - 1,080,567n + 846,720\}. \quad (17)$$

It may be noted that, in terms of κ_2, we have, to the highest order in n, for τ,

$$\kappa_4 = -\frac{54}{25n}\kappa_2^2 \sim -2{\cdot}16\kappa_2^2/n, \quad (18)$$

$$\kappa_6 = \frac{1,296}{49n^2}\kappa_2^3 \sim 26{\cdot}45\kappa_2^3/n^2, \quad (19)$$

$$\kappa_8 = -\frac{3,888}{5n^3}\kappa_2^4 \sim -777{\cdot}6\kappa_2^4/n^3, \quad (20)$$

indicating a fairly rapid approach to normality. For ρ_s we have

$$\kappa_4 = -\frac{114}{25n}\kappa_2^2 \sim -4{\cdot}56\kappa_2^2/n, \quad (21)$$

$$\kappa_6 = \frac{27,984}{245n^2}\kappa_2^3 \sim -114{\cdot}22\kappa_2^3/n^2, \quad (22)$$

$$\kappa_8 = -\frac{6,039,216}{875n^3}\kappa_2^4 \sim -6,902\kappa_2^4/n^3, \quad (23)$$

showing that the tendency is a good deal slower than for τ.

8. The values of the standardized cumulants for $n = 10$ are as follows:

	ρ_s	τ
κ_2	0·111,111	0·061,728,4
κ_4/κ_2^2	−0·464	−0·216,090
κ_6/κ_2^3	1·139,526	0·257,254
κ_8/κ_2^4	−6·560,788	−0·732,829

As far as terms of order n^{-3} the Edgeworth expansion of a standardized symmetrical frequency function is

$$\exp\left\{\frac{\kappa_4 D^4}{4!} + \frac{\kappa_6 D^6}{6!} + \frac{\kappa_8 D^8}{8!}\right\}\alpha(x), \quad (24)$$

where $D = \frac{d}{dx}$ and $\alpha(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$, and this becomes

$$f(x) = \left\{1 + \frac{\kappa_4}{24} D^4 + \left(\frac{\kappa_6}{720} D^6 + \frac{\kappa_4^2}{1{,}152} D^8\right) + \left(\frac{\kappa_8}{40{,}320} D^8 + \frac{\kappa_6 \kappa_4}{17{,}280} D^{10} + \frac{\kappa_4^3 D^{12}}{82{,}944}\right)\right\} \alpha(x)$$

$$= \alpha(x)\left\{1 + \frac{\kappa_4}{24} H_4 + \left(\frac{\kappa_6}{720} H_6 + \frac{\kappa_4^2}{1{,}152} H_8\right) + \left(\frac{\kappa_8}{40{,}320} H_8 + \frac{\kappa_6 \kappa_4}{17{,}280} H_{10} + \frac{\kappa_4^3}{82{,}944} H_{12}\right)\right\}, \quad (25)$$

where H_r is the Tchebycheff-Hermite polynomial defined by

$$(-D)^r \alpha(x) = H_r(x)\,\alpha(x).$$

The distribution function, which is more to the point for our purpose, is given by

$$F(x) = \int_{-\infty}^{x} \alpha(x)\,dx - \alpha(x)\left\{\frac{\kappa_4}{24} H_3 + \left(\frac{\kappa_6}{720} H_5 + \frac{\kappa_4^2}{1{,}152} H_7\right) + \left(\frac{\kappa_8}{40{,}320} H_7 + \frac{\kappa_6 \kappa_4}{17{,}280} H_9 + \frac{\kappa_4^3}{82{,}944} H_{11}\right)\right\}. \quad (26)$$

We shall need to consider values of x not greater than about 3. The explicit expressions for the H's we require are

$$\left.\begin{aligned} H_3 &= x^3 - 3x, \\ H_5 &= x^5 - 10x^3 + 15x, \\ H_7 &= x^7 - 21x^5 + 105x^3 - 105x, \\ H_9 &= x^9 - 36x^7 + 378x^5 - 1{,}260x^3 + 945x, \\ H_{11} &= x^{11} - 55x^9 + 990x^7 - 6{,}930x^5 + 17{,}325x^3 - 10{,}395x. \end{aligned}\right\} \quad (27)$$

9. We now compare the values of the distribution function given by (26) with the exact values for $n = 10$ in the neighbourhood of the significance points 5, 1 and 0·1 %.

A preliminary remark is necessary concerning corrections for continuity. In the usual test for the score S in the τ-distribution the deviate is standardized by division by the standard deviation $\sqrt{\{\frac{1}{18} n(n-1)(2n+5)\}}$ after being reduced by unity (cf. Kendall, 1948, p. 41). We improve on this by correcting the variance by a Sheppard correction, and similarly correct the other cumulants required in the computations. Thus, for $n = 10$, the variance of S is 125. With a Sheppard correction this becomes 124·6667. The deviate appropriate to a value of S of 19 is then $18/\sqrt{124{\cdot}6667} = 1{\cdot}612$. The position is similar for Spearman's ρ. The (corrected) cumulants of ρ_s and τ are:

	ρ_s	τ
κ_4/κ_2^2	−0·4641	−0·2172
κ_6/κ_2^3	1·1399	0·2593
κ_8/κ_2^4	−6·5636	−0·7406

These are the values used in applying equation (26), but, in fact, apart from the initial term, the uncorrected values of § 8 would have given the same final numerical results.

10. Consider τ in the first place. For the complement of the distribution function of equation (26) we have

$$1-F_\tau(x)=\int_x^\infty \alpha(x)\,dx+\alpha(x)\{-0{\cdot}0^2905H_3+0{\cdot}0^336H_5 \\ +0{\cdot}0^44H_7-0{\cdot}0^418H_9-0{\cdot}0^53H_9-0{\cdot}0^61H_{11}\}. \quad (28)$$

The following Table 3 shows the exact values of the probability, the deviate x_0 and the successive approximations P_1, P_2, P_3 and P_4 obtained by using one, two, four and seven terms in (28), i.e. by taking terms of order n^0, n^{-1}, n^{-2} and n^{-3} respectively.

Table 3. *Approximations to the distribution function of* τ *(case,* $n = 10$*)*

S	x_0	P_1	P_2	P_3	P_4	Exact
19	1·612	0·05348	0·05412	0·05416	0·05417	0·05416
21	1·791	0·03665	0·03638	0·03632	0·03630	0·03628
25	2·149	0·01582	0·01457	0·01438	0·01435	0·01430
27	2·329	$0{\cdot}0^2993$	$0{\cdot}0^2858$	$0{\cdot}0^2838$	$0{\cdot}0^2834$	$0{\cdot}0^2833$
33	2·865	$0{\cdot}0^22085$	$0{\cdot}0^21196$	$0{\cdot}0^21117$	$0{\cdot}0^21110$	$0{\cdot}0^21106$
35	3·045	$0{\cdot}0^21160$	$0{\cdot}0^3491$	$0{\cdot}0^3459$	$0{\cdot}0^3464$	$0{\cdot}0^3473$

The approximation P_4 is very good, considering that n is only 10 and that we are examining the tails of a discontinuous distribution. Even P_2 is very fair and would not lead to incorrect inferences in practice. For larger values of n the approximations would doubtless be closer.

11. Similar calculations were carried out for ρ_s for $n = 10$ with the results shown in Table 4. The series corresponding to (28) was

$$1-F_\rho(x)=\int_x^\infty \alpha(x)\,dx+\alpha(x)\{-0{\cdot}01933H_3+0{\cdot}0^2158H_5+0{\cdot}0^3187H_7 \\ -0{\cdot}0^3163H_7-0{\cdot}0^43H_9-0{\cdot}0^512H_{11}\}. \quad (29)$$

Table 4. *Approximations to the distribution function of* ρ_s *(case,* $n = 10$*)*

$S(d^2)$	x_0	P_1	P_2	P_3	P_4	Exact
256	1·636	0·05094	0·05201	0·05219	0·05225	0·05244
258	1·673	0·04719	0·04783	0·04792	0·04794	0·04814
286	2·182	0·01456	0·01182	0·01096	0·01064	0·01012
288	2·218	0·01327	0·01046	$0{\cdot}0^2959$	$0{\cdot}0^2926$	$0{\cdot}0^2870$
308	2·582	$0{\cdot}0^2491$	$0{\cdot}0^2230$	$0{\cdot}0^2160$	$0{\cdot}0^2135$	$0{\cdot}0^2109$
310	2·618	$0{\cdot}0^2442$	$0{\cdot}0^3189$	$0{\cdot}0^2123$	$0{\cdot}0^3998$	$0{\cdot}0^3800$

It was to be expected that the agreement for ρ_s would be less satisfactory than for τ. Nevertheless, the approximation represented by P_4 is good and would be quite satisfactory in practice. In fact, since $S(d^2)$ is discontinuous P_4 would indicate quite correctly which pair of values bracketed the significance levels. When it is remembered that we are dealing with the tails of a discontinuous distribution of finite range, the approximation seems remarkably good.

THE RELATIONSHIP BETWEEN r_s AND t IN NORMAL SAMPLES

12. In the null case the distribution of r_s and t tends to bivariate normality for large n with a correlation which fairly rapidly tends to unity. On the basis of this result and the known

expressions for the variances of r_s and t in the null case, Kendall has remarked that for moderate values of parental correlation r_s in practice may be expected to be about $1{\cdot}5t$. Daniels (1950) has, however, quite rightly pointed out that it may be unjustifiable to generalize from the null case in this way if the parent is heterogeneous. He proves the interesting inequality for large n

$$-1 \leqslant 3t - 2r_s \leqslant 1, \tag{30}$$

and shows that if $t > 0$ the upper limit can be attained but not the lower limit, if $t < 0$ the reverse is true and if $t = 0$ both limits are attainable. He goes on to remark that in certain cases it is possible to have a population for which $\tau = 0$ but $\rho_s = \pm\frac{1}{2}$, and hence that τ and ρ_s may be measuring different aspects of the population.

13. That such a case is possible is undeniable; but in what might be called 'respectable' variation, namely, when the parent is capable of representation by measurable variates which are fairly close to joint normality, the correlation between r_s and t remains quite high. We shall establish this result by extending the method used by Kendall (1949) to obtain the variance of r_s in normal samples. Incidentally, we extend his expansion as far as powers of ρ^{10} and correct a small error in his coefficient of ρ^8.

In the usual notation we have

$$t = \frac{1}{n(n-1)} \Sigma a_{ij} b_{ij}, \tag{31}$$

$$r_s = \frac{3}{n(n^2-1)} \sum_{i,j} a_{ij} \sum_k b_{ik}, \tag{32}$$

and in normal samples

$$E(t) = E(a_{ij} b_{ij}) = \frac{2}{\pi} \sin^{-1} \rho, \tag{33}$$

$$E(a_{ij} b_{ik}) = \frac{2}{\pi} \sin^{-1} \tfrac{1}{2}\rho, \tag{34}$$

$$E(r_s) = \frac{6}{\pi(n+1)} \{\sin^{-1} \rho + (n-2) \sin^{-1} \tfrac{1}{2}\rho\}. \tag{35}$$

For the variance of r_s we have to order n^{-1}

$$\begin{aligned} \operatorname{var} r_s = \frac{9}{n} [&- 9\{E^2(a_{ij} b_{ik})\} + E(a_{ij} b_{ik} a_{im} b_{ip}) \\ &+ 4E(a_{ij} b_{ij} a_{im} b_{ip}) + 2E(a_{ij} b_{ik} a_{lm} b_{lk}) + 2E(a_{ij} b_{ik} a_{lm} b_{lj})], \end{aligned} \tag{36}$$

and for the variance of t, exactly

$$\operatorname{var} t = \frac{1}{\frac{1}{2}n(n-1)} \left[1 - \left(\frac{2}{\pi} \sin^{-1} \rho \right)^2 + 2(n-2) \left\{ \frac{1}{9} - \left(\frac{2}{\pi} \sin^{-1} \tfrac{1}{2}\rho \right)^2 \right\} \right], \tag{37}$$

or to order n^{-1}

$$\operatorname{var} t = \frac{4}{n} \left\{ \frac{1}{9} - \left(\frac{2}{\pi} \sin^{-1} \tfrac{1}{2}\rho \right)^2 \right\}. \tag{38}$$

We then find, by the method described by Kendall (1949),

$$E(a_{ij} b_{ik}) = \left(\frac{2}{\pi} \sin^{-1} \tfrac{1}{2}\rho \right)^2 = \left(\frac{2}{\pi} \right)^2 \left[\left(\frac{\rho}{2} \right)^2 + \frac{1}{3} \left(\frac{\rho}{2} \right)^4 + \frac{8}{45} \left(\frac{\rho}{2} \right)^6 + \frac{4}{35} \left(\frac{\rho}{2} \right)^8 + \frac{128}{1{,}575} \left(\frac{\rho}{2} \right)^{10} \right], \tag{39}$$

$$E(a_{ij} b_{ik} a_{im} b_{ip}) = \left(\frac{2}{\pi} \right)^2 \left[\frac{\pi^2}{36} + \frac{2}{3} \left(\frac{\rho}{2} \right)^2 + \frac{50}{81} \left(\frac{\rho}{2} \right)^4 + \frac{784}{1{,}215} \left(\frac{\rho}{2} \right)^6 + \frac{55{,}112}{76{,}545} \left(\frac{\rho}{2} \right)^8 + \frac{26{,}050{,}816}{31{,}000{,}725} \left(\frac{\rho}{2} \right)^{10} \right], \tag{40}$$

$$E(a_{ij}b_{il}a_{lm}b_{lp}) = \left(\frac{2}{\pi}\right)^2\sqrt{\frac{4}{3}}\left[\frac{1}{2}\left(\frac{\rho}{2}\right)^2+\frac{2}{9}\left(\frac{\rho}{2}\right)^4+\frac{23}{135}\left(\frac{\rho}{2}\right)^6+\frac{52}{315}\left(\frac{\rho}{2}\right)^8+\frac{23{,}341}{127{,}575}\left(\frac{\rho}{2}\right)^{10}\right], \quad (41)$$

$$E(a_{ij}b_{ik}a_{lm}b_{lk}) = 2E(a_{ij}b_{il}a_{lm}b_{lp}), \quad (42)$$

$$E(a_{ij}b_{ik}a_{lm}b_{lj}) = \left(\frac{2}{\pi}\right)^2\left[\left(\frac{\rho}{2}\right)^2+\frac{5}{6}\left(\frac{\rho}{2}\right)^2+\frac{379}{360}\left(\frac{\rho}{2}\right)^6+\frac{869}{560}\left(\frac{\rho}{2}\right)^8+\frac{222{,}287}{100{,}800}\left(\frac{\rho}{2}\right)^{10}\right], \quad (43)$$

whence, on substitution in (36),

$$\operatorname{var} r_s = \frac{1}{n}(1-1{\cdot}563{,}464{,}96\rho^2+0{\cdot}304{,}742{,}82\rho^4+0{\cdot}155{,}285{,}69\rho^6 + 0{\cdot}061{,}551{,}66\rho^8+0{\cdot}022{,}099{,}22\rho^{10}). \quad (44)$$

By expanding (38) we find also

$$\operatorname{var} t = \frac{4}{9n}(1-0{\cdot}911{,}890{,}65\rho^2-0{\cdot}075{,}990{,}89\rho^4-0{\cdot}010{,}132{,}12\rho^6 - 0{\cdot}001{,}628{,}37\rho^8-0{\cdot}000{,}289{,}49\rho^{10}). \quad (45)$$

14. We now evaluate the covariance. We have

$$\operatorname{cov}(r_s,t) = E(r_s t)-E(r_s)E(t) \quad (46)$$

and

$$E(r_s t) = E\frac{\sum_{i,j}(a_{ij}b_{ij})\sum_{i,j}\sum a_{ij}\sum_k b_{ik}}{\frac{1}{3}n^2(n-1)^2(n+1)},$$

which is equivalent to

$$\tfrac{1}{3}n^2(n-1)^2(n+1)E(r_s t) = E\Sigma a_{ij}b_{ij}a_{mk}b_{ml},$$

where the suffices i, j, m, k, l all range from 1 to n. To order n^4 the right-hand side gives terms of four types, namely,

$$n^{(5)}E(a_{ij}b_{ij}a_{mk}b_{ml})+n^{(4)}E(a_{ij}b_{ij}a_{mk}b_{mk})+4n^{(4)}E(a_{ij}b_{ij}a_{mk}b_{mi})+2n^{(4)}E(a_{ij}b_{ij}a_{ik}b_{il}),$$

where suffices with different letters are now different. Thus, to order n^{-1},

$$E(r_s t) = \frac{3}{n}[(n-9)E(a_{ij}b_{ij}a_{mk}b_{ml})+E(a_{ij}b_{ij}a_{mk}b_{mk})+4E(a_{ij}b_{ij}a_{mk}b_{mi})+2E(a_{ij}b_{ij}a_{ik}b_{il})]. \quad (47)$$

On substitution in (46) we find to order n^{-1}

$$\operatorname{cov}(r_s t) = \frac{3}{n}[4E(a_{ij}b_{ij}a_{mk}b_{mi})+2E(a_{ij}b_{ij}a_{ik}b_{il})-6E(a_{ij}b_{ij})E(a_{mk}b_{ml})]. \quad (48)$$

It remains to evaluate these expectations. In the manner followed by Kendall (1949) we find

$$E(a_{ij}b_{ij}a_{mk}b_{mi}) = \frac{1}{\pi^2\sqrt{3}}\left[12\left(\frac{\rho}{2}\right)^2+\frac{40}{3}\left(\frac{\rho}{2}\right)^4+\frac{368}{15}\left(\frac{\rho}{2}\right)^6+\frac{17{,}824}{315}\left(\frac{\rho}{2}\right)^8+\frac{18{,}267{,}584}{127{,}575}\left(\frac{\rho}{2}\right)^{10}\right], \quad (49)$$

$$E(a_{ij}b_{ij}a_{ik}b_{il}) = \frac{1}{\pi^2}\left[\frac{\pi^2}{9}+\frac{14}{3}\left(\frac{\rho}{2}\right)^2+\frac{470}{81}\left(\frac{\rho}{2}\right)^4+\frac{12{,}208}{1{,}215}\left(\frac{\rho}{2}\right)^6 + \frac{1{,}653{,}848}{76{,}545}\left(\frac{\rho}{2}\right)^8+\frac{1{,}680{,}524{,}032}{31{,}000{,}725}\left(\frac{\rho}{2}\right)^{10}\right], \quad (50)$$

Hence, on substitution in (48), we find

$$\operatorname{cov}(r_s,t) = \frac{2}{3n}(1-1{\cdot}248{,}589{,}61\rho^2+0{\cdot}068{,}304{,}96\rho^4+0{\cdot}072{,}804{,}83\rho^6 + 0{\cdot}040{,}255{,}26\rho^8+0{\cdot}016{,}413{,}62\rho^{10}). \quad (51)$$

To evaluate the correlation between r_s and t it is tempting to expand $\operatorname{cov}(r_s t)/\sqrt{(\operatorname{var} r_s \operatorname{var} t)}$ as a power series in ρ. The convergence of such an expansion, however, is less satisfactory than that of the three series for $\operatorname{cov}(r_s t)$, $\operatorname{var} r_s$ and $\operatorname{var} t$ separately, and it is best to evaluate them directly. The results for a range of values of ρ from 0 to 0·9 are shown in Table 5.

Table 5. *Correlation between r_s and t in normal samples for large n*

ρ	var r_s from (44) $1/n \times$	var t from (45) $4/9n \times$	cov $r_s t$ from (51) $2/3n \times$	correlation, r_s and t
0·0	1·000,000	1·000,000	1·000,000	1·000,000
0·1	0·984,396	0·990,873	0·987,521	0·999,89
0·2	0·937,959	0·963,402	0·950,170	0·999,55
0·3	0·861,874	0·917,307	0·888,236	0·998,96
0·4	0·758,326	0·852,109	0·802,301	0·998,07
0·5	0·630,869	0·767,123	0·693,433	0·996,79
0·6	0·485,060	0·661,369	0·563,490	0·994,87
0·7	0·329,512	0·533,688	0·415,941	0·991,86
0·8	0·177,612	0·382,303	0·256,482	0·984,28
0·9	0·050,261,8	0·205,320	0·095,200,4	0·937,14

The correlation, as will be seen, is very high until ρ itself approaches unity. In the neighbourhood of $\rho = 1$ the covariance of r_s and t tends to zero (because their variances tend to zero), but each coefficient, of course, will tend to unity. For large n, therefore, the ratios of the coefficients will not differ very much from the ratio of their expectations, namely, $3\sin^{-1}\frac{1}{2}\rho/\sin^{-1}\rho$, which varies from 1·5 in the neighbourhood of $\rho = 0$ to unity when $\rho = 1$. When $\rho = 0{\cdot}6$ the ratio is 1·42.

REFERENCES

Daniels, H. E. (1950). *J. R. Statist. Soc.* B, **12**, 171.
David, F. N. & Kendall, M. G. (1949). *Biometrika*, **36**, 431.
Hotelling, H. & Pabst, M. R. (1936). *Ann. Math. Statist.* **7**, 29.
Kendall, M. G. (1948). *Rank Correlation Methods.* London: Charles Griffin and Co.
Kendall, M. G. (1949). *Biometrika*, **36**, 177.
Kendall, M. G. *et al.* (1939). *Biometrika*, **30**, 251.
Olds, E. G. (1938). *Ann. Math. Statist.* **9**, 133.
Moran, P. A. (1950). *J. R. Statist. Soc.* B, **12**, 153.
Silverstone, H. (1950). *Biometrika*, **37**, 231.
Whitfield, J. W. (1950). *J. R. Statist. Soc.* B, **12**, 163.

NOTE ON AN EXACT TREATMENT OF CONTINGENCY, GOODNESS OF FIT AND OTHER PROBLEMS OF SIGNIFICANCE

By G. H. FREEMAN and J. H. HALTON

I. The present treatment was evolved because of the inaccuracy of the use of χ^2 tables when expected and observed numbers are small. The method is exact and applies to any sampled array of numbers, whatever the population from which the sampling is effected, provided that:

(*a*) either the parent population is infinite, or the sampling is done with replacement of the sampled members;

(*b*) the sampling is random.

The treatment may also be used to determine the significance of effects whose nature is such that Pearson's χ^2 test is not applicable.

The only difficulty encountered in the use of the method outlined below is the amount of labour involved in obtaining the results, and this sets an upper limit to the size of the sample that can be dealt with in a reasonable time, using hand-operated or semi-automatic machines.

II. Consider any population whose members may be subjected to a complete and exclusive classification into a finite number, r, of groups. Then the probability that out of a random sample of n members, l_i will fall into the ith group ($i = 1, 2, \ldots, r$) is

$$P_L = n! \prod_{i=1}^{r} \left(\frac{p_i^{l_i}}{l_i!}\right),$$

where p_i is the probability that a randomly selected member will belong to the ith group. Putting $\lambda_i = np_i$, this becomes

$$P_L = \left(\frac{n!}{n^n}\right) \prod_{i=1}^{r} \left(\frac{\lambda_i^{l_i}}{l_i!}\right), \tag{1}$$

where λ_i is the number of members of the sample expected to fall into the ith group.

In addition, should the choice of the sample considered, although otherwise random, be restricted by a number of conditions $C_1, C_2, \ldots, C_s$, whose *a priori* probabilities are respectively $P_{C_1}, P_{C_2}, \ldots, P_{C_s}$, the expression (1) must be divided by their product, giving

$$P_L = \left(\frac{n!}{n^n}\right) \prod_{i=1}^{r} \left(\frac{\lambda_i^{l_i}}{l_i!}\right) \Big/ \prod_{j=1}^{s} (P_{C_j}). \tag{2}$$

III. Similarly, if a population is subjected to k different and independent classifications, the corresponding probability of obtaining an array $L(l_{i_1 i_2 \ldots i_k})$ by random sampling is

$$P_L = \left(\frac{n!}{n^n}\right) \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \cdots \prod_{i_k=1}^{r_k} \left(\frac{\lambda_{i_1 i_2 \ldots i_k}^{l_{i_1 i_2 \ldots i_k}}}{l_{i_1 i_2 \ldots i_k}!}\right) \Big/ \prod_{j=1}^{s} (P_{C_j}). \tag{3}$$

For most purposes, only relationships internal to the contingency table formed for the sample considered are of interest. In such cases, the numbers of members of the sample in the ith group of the mth classification ($i_m = 1, 2, \ldots, r_m$; $m = 1, 2, \ldots, k$), i.e. the 'border totals', may be taken as fixed. If this is done, and if further:

$$a_{mi_m} = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \cdots \sum_{i_{m-1}=1}^{r_{m-1}} \sum_{i_{m+1}=1}^{r_{m+1}} \cdots \sum_{i_k=1}^{r_k} (l_{i_1 i_2 \ldots i_k}),$$

it is found, on substituting the *a priori* probability of obtaining arrays $A_m(a_{mi_m})$ for the P_{C_j} in (3) that

$$P_L = \left(\frac{n^n}{n!}\right)^{k-1} \left\{ \frac{\prod_{m=1}^{k} \prod_{i_m=1}^{r_m} (a_{mi_m}!)}{\prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \cdots \prod_{i_k=1}^{r_k} (l_{i_1 i_2 \ldots i_k}!)} \right\} \left\{ \frac{\prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \cdots \prod_{i_k=1}^{r_k} (\lambda_{i_1 i_2 \ldots i_k}^{l_{i_1 i_2 \ldots i_k}})}{\prod_{m=1}^{k} \prod_{i_m=1}^{r_m} (a_{mi_m}^{a_{mi_m}})} \right\}. \tag{4}$$

Finally, if it is assumed that the population is homogeneous, an assumption which is equivalent to

$$\lambda_{i_1 i_2 \ldots i_k} = \frac{\prod_{m=1}^{k} (a_{mi_m})}{n^{k-1}}, \tag{5}$$

then, on substitution into (4), it is found that

$$P_L = \frac{\prod_{m=1}^{k} \prod_{i_m=1}^{r_m} (a_{mi_m}!)}{(n!)^{k-1} \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \cdots \prod_{i_k=1}^{r_k} (l_{i_1 i_2 \ldots i_k}!)}. \tag{6}$$

IV. A special case of importance is that in which $k = 2$. Putting i for i_1, j for i_2, r for r_1, s for r_2, a_i for a_{1i_1} and b_j for a_{2i_2}, the expressions (4) and (6) become respectively

$$P_L = \left(\frac{n^n}{n!}\right) \left\{ \frac{\prod_{i=1}^{r} (a_i!) \prod_{j=1}^{s} (b_j!)}{\prod_{i=1}^{r} \prod_{j=1}^{s} (l_{ij}!)} \right\} \left\{ \frac{\prod_{i=1}^{r} \prod_{j=1}^{s} (\lambda_{ij}^{l_{ij}})}{\prod_{i=1}^{r} (a_i^{a_i}) \prod_{j=1}^{s} (b_j^{b_j})} \right\}, \tag{7}$$

and

$$P_L = \frac{\prod_{i=1}^{r} (a_i!) \prod_{j=1}^{s} (b_j!)}{n! \prod_{i=1}^{r} \prod_{j=1}^{s} (l_{ij}!)}. \tag{8}$$

When $r = s = 2$ we get the known result for the exact treatment of 2×2 tables (see, for instance, R. A. Fisher's *Statistical Methods for Research Workers*, §21·02).

Further, the k-variate contingency table, when assumed to be homogeneous, may be looked upon from the point of view of the heterogeneity due to one classification only, say the hth. In that case, the contingency table may be looked upon as a bivariate contingency table, the two classifications being the hth classification on the one hand, and the $(k-1)$ other classifications on the other hand.

Putting

$$d_{i_1 i_2 \ldots i_{h-1} i_{h+1} \ldots i_k} = \sum_{i_h=1}^{r_h} (l_{i_1 i_2 \ldots i_k}),$$

it will be seen that

$$P_L = \frac{\prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \cdots \prod_{i_{h-1}=1}^{r_{h-1}} \prod_{i_{h+1}=1}^{r_{h+1}} \cdots \prod_{i_k=1}^{r_k} (d_{i_1 i_2 \ldots i_{h-1} i_{h+1} \ldots i_k}!) \prod_{i_h=1}^{r_h} (a_{hi_h}!)}{n! \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} \cdots \prod_{i_k=1}^{r_k} (l_{i_1 i_2 \ldots i_k}!)}. \tag{9}$$

V. The actual test is carried out as follows:

(i) All arrays $X^{(\omega)}(x_{i_1 i_2 \ldots i_k}^{(\omega)})$, subject to the same general conditions as the observed array $L(l_{i_1 i_2 \ldots i_k})$, are written out.

(ii) The corresponding *a priori* probabilities $P_{X^{(\omega)}}$ are calculated by means of the appropriate probability expression.

(iii) The values of $P_{X^{(\omega)}}$ satisfying the following equation are noted:

$$P_{X^{(\omega)}} \leqslant P_L. \tag{10}$$

These are the probabilities of all arrays $X^{(t)}$ which are *a priori* as probable as, or less probable than, the array L.

(iv) The probabilities satisfying (10), which include P_L itself, are added together to give a probability $\wp_L$, the summation being over values of t satisfying (10), thus

$$\wp_L = \sum_t (P_{X^{(t)}}). \tag{11}$$

It will be seen that the resulting $\wp_L$ is the total *a priori* probability of obtaining an array as probable (individually) as, or less probable than, the observed array L.

Now $\wp_L$ satisfies the following conditions:

(*a*) $\wp_L$ has one and only one value for every array L;

(*b*) $\wp_L$ has its maximum value, unity, when L is as close to the expected array Λ as is allowed by the condition that the observed numbers are integers;

(*c*) $\wp_L > \wp_{L'}$, if (but not only if) every observed number in L is as close to the corresponding expected number in Λ as is the corresponding number in L', or closer.

We may thus conclude that the magnitude of $\wp_L$ is a measure of the goodness of fit of the observed array L to the expected array Λ.

The χ^2 tables normally used involve the assumption that

$$P_L = Ce^{-\frac{1}{2}(\chi^2)_L}. \tag{12}$$

(where C is a constant and $(\chi^2)_L$ is the 'χ^2' calculated for the array L), which holds with increasing accuracy as observed and expected numbers become larger. Now, all arrays such that (10) holds are arrays such that

$$(\chi^2)_{X^{(t)}} \geqslant (\chi^2)_L. \tag{13}$$

Thus the probability obtained from χ^2 tables is an approximation to $\wp_L$ which becomes closer as observed and expected numbers become larger.

VI. When the test is carried out in practice it is found that none of the operations mentioned in §V above is very lengthy, as the following two examples show. The working of these examples is done more fully than would be the case in practice, in order to clarify the exposition. The first example is of a (2×3) table and the second of a $(2 \times 2 \times 2)$ table, both being assumed homogeneous.

(i) The sample consists of fifty candidates who had passed through a training establishment having previously passed successfully through a selection centre. The candidates are classified by selection centre grade A, B, or C, by training establishment grade X or Y and also by an educational factor M or N. (Totals are given in parentheses, totals of totals in brackets and the overall figure is in black.)

	C	*B*	*A*			*C*	*B*	*A*			*C*	*B*	*A*	
X	0	3	2	(5)	*X*	8	4	0	(12)	*X*	(8)	(7)	(2)	[17]
Y	6	5	1	(12)	*Y*	16	5	0	(21)	*Y*	(22)	(10)	(1)	[33]
	(6)	(8)	(3)	[17]		(24)	(9)	(0)	[33]		[30]	[17]	[3]	50
Education *M*					Education *N*					Overall				

The first step is to test each of the above three bivariate distributions for heterogeneity. This is done by means of the χ^2 test in the second and third tables, in the latter pooling the A and B groups as this has little effect on the heterogeneity. The corresponding values of χ^2 are 0·3492 (D.F. = 1) and 1·7974 (D.F. = 1), respectively, giving $0{\cdot}50 < P < 0{\cdot}70$ and $0{\cdot}10 < P < 0{\cdot}20$.

In the first table no pooling will make the χ^2 tables reliable as the numbers are too small and so the present method is resorted to. The table is assumed to be homogeneous, so that expression (8) is the appropriate one. The procedure is described below and follows the lines indicated in §V. The first step is to write out all the arrays subject to the same general conditions as L, the observed array. In following the method used in this example it is best to do this in full, but in the alternative method used for the second example here this is unnecessary, for reasons given in §VI (ii) below. The arrays, 18 in number, of which the second is the observed array L, are written out as follows:

$$X^{(1)}\begin{pmatrix}0&2&3\\6&6&0\end{pmatrix}, \quad L = X^{(2)}\begin{pmatrix}0&3&2\\6&5&1\end{pmatrix}, \quad X^{(3)}\begin{pmatrix}0&4&1\\6&4&2\end{pmatrix}, \quad X^{(4)}\begin{pmatrix}0&5&0\\6&3&3\end{pmatrix},$$

$$X^{(5)}\begin{pmatrix}1&1&3\\5&7&0\end{pmatrix}, \quad X^{(6)}\begin{pmatrix}1&2&2\\5&6&1\end{pmatrix}, \quad X^{(7)}\begin{pmatrix}1&3&1\\5&5&2\end{pmatrix}, \quad X^{(8)}\begin{pmatrix}1&4&0\\5&4&3\end{pmatrix},$$

$$X^{(9)}\begin{pmatrix}2&0&3\\4&8&0\end{pmatrix}, \quad X^{(10)}\begin{pmatrix}2&1&2\\4&7&1\end{pmatrix}, \quad X^{(11)}\begin{pmatrix}2&2&1\\4&6&2\end{pmatrix}, \quad X^{(12)}\begin{pmatrix}2&3&0\\4&5&3\end{pmatrix},$$

$$X^{(13)}\begin{pmatrix}3&0&2\\3&8&1\end{pmatrix}, \quad X^{(14)}\begin{pmatrix}3&1&1\\3&7&2\end{pmatrix}, \quad X^{(15)}\begin{pmatrix}3&2&0\\3&6&3\end{pmatrix}, \quad X^{(16)}\begin{pmatrix}4&0&1\\2&8&2\end{pmatrix},$$

$$X^{(17)}\begin{pmatrix}4&1&0\\2&7&3\end{pmatrix}, \quad X^{(18)}\begin{pmatrix}5&0&0\\1&8&3\end{pmatrix}.$$

The ordering of the arrays above is done on the following principle. In any $(r_1 \times r_2 \times \ldots \times r_k)$ contingency table in which only the border totals, and hence the overall total, are fixed there are f degrees of freedom where

$$f = \prod_{m=1}^{k} (r_m) - \sum_{m=1}^{k} (r_m - 1) - 1.$$

f independent cells are then picked out and placed in arbitrary order. The ordered set of f numbers obtained is then regarded as an f-figure 'number' and all possible 'numbers' are written out in order of size. In this example the figures in the CX and BX cells, in that order, form the 'number', f being 2.

The second step is the calculation of the separate probabilities $P_{X^{(t)}}$, and this is done here by breaking down each $P_{X^{(t)}}$ into two factors Q_L and $R_{X^{(t)}}$, where Q_L is independent of t.

Thus
$$P_{X^{(t)}} = Q_L \div R_{X^{(t)}},$$

and since
$$P_{X^{(t)}} = \frac{\prod_{i=1}^{r} (a_i!) \prod_{j=1}^{s} (b_j!)}{n! \prod_{i=1}^{r} \prod_{j=1}^{s} (x_{ij}^{(t)}!)},$$

we have
$$Q_L = \frac{\prod_{i=1}^{r} (a_i!) \prod_{j=1}^{s} (b_j!)}{n!}$$

and $$R_{X^{(t)}} = \prod_{i=1}^{r} \prod_{j=1}^{s} (x_{ij}^{(t)}!).$$

By the use of logarithms the calculations are then performed, Q_L being determined once for all and each $R_{X^{(t)}}$ separately. In this example the values of $P_{X^{(t)}}$ are then found as follows, that of $P_L = P_{X^{(0)}}$ being given in black.

t	1	2	3	4	5	6	7	8	9
$P_{X^{(t)}}$	0·0045	0·0271	0·0339	0·0090	0·0078	0·0815	0·1629	0·0679	0·0024
t	10	11	12	13	14	15	16	17	18
$P_{X^{(t)}}$	0·0582	0·2037	0·1357	0·0097	0·0776	0·0905	0·0073	0·0194	0·0010

As a check, all the $P_{X^{(t)}}$ are added together to give $\sum_{t=1}^{18} P_{X^{(t)}} = 1{\cdot}0001$, the error being easily covered by the corrections to four decimal places in the working.

The values of t satisfying inequality (10) are then noted, these being $t = 1, 2, 4, 5, 9, 13, 16, 17, 18$. For these values of t the values of $P_{X^{(t)}}$ are summed to give $\wp_L$.

Thus $\wp_L = 0{\cdot}0882$ and so, at the 5 % level, there is no evidence of heterogeneity in the distribution of selection centre and training establishment grades in the case of candidates with educational factor M, in the sample considered.

Since in the case of these candidates the probability $\wp_L$ is not far removed from the significance level 0·05, we may test whether there is evidence of positive association between the two sets of grades in this sample, regarding X as better than Y, and A as better than B, B being better than C. Of the arrays set out above, those with positive regression coefficients are those where $t = 1$ to 11 and 13. Of these, those with $t = 1, 2, 4, 5, 9$ and 13 satisfy inequality (10).

Note. For a general (2×3) bivariate table, with the arbitrary values shown here, the regression coefficients are positive if:

$$(c+d)-(a+f) > \frac{(E-C)(A-B)}{N^2}.$$

a d	b e	c f	A B	$+1$ -1
C	D	E	N	
-1	0	$+1$		

In the case of the array considered above, the right-hand side of this inequality has the value $\frac{(-3)(-7)}{17^2}$, or 0·0727. Thus the arrays with positive regression coefficients are those for which $(c+d)-(a+f) \geqslant 1$.

Considering only arrays with positive regression coefficients we have, writing P_i for $P_{X^{(i)}}$:

$$P_{L,\text{positive}} = \frac{P_1+P_2+P_4+P_5+P_9+P_{13}}{\sum_{i=1}^{11} P_i + P_{13}} = \frac{0{\cdot}0605}{0{\cdot}6686} = 0{\cdot}0905.$$

Thus there is no significant evidence of association between the grades given to candidates with educational factor M at the selection centre and at the training establishment, in this sample.

This test for evidence of association is one which cannot be performed by means of χ^2 but which can be readily done by the present method.

(ii) The second example considered consists of the attainments of a set of 25 children. These are classified according to year of birth (1935/6 and 1937/8/9), attendances at school during a particular fortnight (up to 7, and 8 or more) and possession of a certain certificate or not. The distribution of the classifications is given below, totals as before being in parentheses, totals of totals in brackets and the overall total being in black.

Total number of attendances											
Up to 7				8 or more				Total			
	Year of birth				Year of birth				Year of birth		
	1935/6	1937/8/9			1935/6	1937/8/9			1935/6	1937/8/9	
Cert.	0	0	(0)	Cert.	1	2	(3)	Cert.	(1)	(2)	[3]
Not	1	4	(5)	Not	1	16	(17)	Not	(2)	(20)	[22]
	(1)	(4)	[5]		(2)	(18)	[20]		[3]	[22]	25

The array is considered as a $(2 \times 2 \times 2)$ trivariate contingency table. The total number of children in each group of each classification is assumed to be fixed and the array tested for general homogeneity or heterogeneity, the appropriate expression being expression (6) with $k = 3, r_k = 2$. In the method described in this example not all the arrays $X^{(i)}$ obtained as before need to be written out but only certain key ones, on principles given below.

On the same system of ordering as used in the first example, $f = 4$ and the four cells the numbers in which form the four-figure 'number' are all but the bottom right-hand cell in the first table above, first the top row from left to right, then the bottom left-hand one, and finally the top left-hand cell of the second table. On this system $X^{(1)}$ and $X^{(2)}$ are respectively:

$$X^{(1)}\left(\begin{array}{cc|cc} 0 & 0 & 0 & 3 \\ 0 & 5 & 3 & 14 \end{array}\right) \quad \text{and} \quad X^{(2)}\left(\begin{array}{cc|cc} 0 & 0 & 1 & 2 \\ 0 & 5 & 2 & 15 \end{array}\right).$$

The expression (6) for the probability may be here written as:

$$P_{X^{(i)}} = \frac{\text{Product of all (border totals!)}}{(n!)^2 \times \text{product of all (cell totals!)}}.$$

On this principle we have:

$$P_{X^{(1)}} = \frac{5!\,20!\,3!\,22!\,3!\,22!}{(25!)^2\,0!\,0!\,0!\,5!\,0!\,3!\,3!\,14!}$$

$$= \frac{6 \times 17 \times 19}{23 \times 23 \times 25},$$

$$P_{X^{(2)}} = \frac{5!\,20!\,3!\,22!\,3!\,22!}{(25!)^2\,0!\,0!\,0!\,5!\,1!\,2!\,2!\,15!}$$

$$= P_{X^{(1)}} \times \frac{0!\,3!\,3!\,14!}{1!\,2!\,2!\,15!} = P_{X^{(1)}} \frac{(3 \times 3)}{(1 \times 15)}.$$

The principle of this last calculation is clear. Given any $P_{X^{(t_1)}}$, if the difference between corresponding cell-totals of $X^{(t_1)}$ and $X^{(t_2)}$ is either one or zero, the value of $P_{X^{(t_2)}}$ is the value of $P_{X^{(t_1)}}$ multiplied by the greater of differing pairs of cell-totals if they occur in $X^{(t_1)}$ and divided by them if they occur in $X^{(t_2)}$.

If $P^{(t_2)} = P^{(t_1)} \times \phi(t_1, t_2)$, we see that the factor $\phi(t_1, t_2)$ is subject to the following conditions:

(i) Every $\phi(t_1, t_2)$ consists of a fraction of which the numerator and denominator contain the same number of integral factors, in general two.

(ii) If the arrays are numbered systematically in the order explained above, then whenever both the pairs of arrays $X^{(t+1)}$ and $X^{(t)}$, and $X^{(t)}$ and $X^{(t-1)}$, satisfy the required conditions for X_{t_1} and X_{t_2} we have:

(a) Each integer in the numerator of $\phi(t-1, t)$ is greater by one than the corresponding integer in the numerator of $\phi(t, t+1)$.

(b) Each integer in the denominator of $\phi(t-1, t)$ is less by one than the corresponding integer in the denominator of $\phi(t, t+1)$.

(iii) If condition (ii) is not satisfied, there is a break in the sequence of factors ϕ.

It will thus be seen that not all the arrays need be written out. If $X^{(1)}$ and $X^{(2)}$ are written out and also those values of $X^{(t)}$ and $X^{(t+1)}$ for which $\phi(t', t)$ does not follow from $\phi(t'', t-1)$, irrespective of t' and t'', which occurs when $\phi(t'', t-1)$ has unity as a factor in its numerator, no other arrays need be written out; for in all other cases the values of $\phi(t, t+1)$ can be deduced from those of the preceding $\phi(t-1, t)$. By this method a substantial saving of time and labour is effected and these rules are followed in writing out the arrays below.

$$L = X^{(1)}\left(\begin{array}{cc|cc} 0 & 0 & 0 & 3 \\ 0 & 5 & 3 & 14 \end{array}\right), \quad X^{(2)}\left(\begin{array}{cc|cc} 0 & 0 & 1 & 2 \\ 0 & 5 & 2 & 15 \end{array}\right), \quad \ldots, \quad X^{(5)}\left(\begin{array}{cc|cc} 0 & 0 & 0 & 3 \\ 1 & 4 & 2 & 15 \end{array}\right),$$

$$X^{(6)}\left(\begin{array}{cc|cc} 0 & 0 & 1 & 2 \\ 1 & 4 & 1 & 16 \end{array}\right), \quad \ldots, \quad X^{(8)}\left(\begin{array}{cc|cc} 0 & 0 & 0 & 3 \\ 2 & 3 & 1 & 16 \end{array}\right), \quad X^{(9)}\left(\begin{array}{cc|cc} 0 & 0 & 1 & 2 \\ 2 & 3 & 0 & 17 \end{array}\right),$$

$$X^{(10)}\left(\begin{array}{cc|cc} 0 & 0 & 0 & 3 \\ 3 & 2 & 0 & 17 \end{array}\right), \quad X^{(11)}\left(\begin{array}{cc|cc} 0 & 1 & 0 & 2 \\ 0 & 4 & 3 & 15 \end{array}\right), \quad X^{(12)}\left(\begin{array}{cc|cc} 0 & 1 & 1 & 1 \\ 0 & 4 & 2 & 16 \end{array}\right), \quad \ldots,$$

$$X^{(14)}\left(\begin{array}{cc|cc} 0 & 1 & 0 & 2 \\ 1 & 3 & 2 & 16 \end{array}\right), \quad X^{(15)}\left(\begin{array}{cc|cc} 0 & 1 & 1 & 1 \\ 1 & 3 & 1 & 17 \end{array}\right), \quad \ldots, \quad X^{(17)}\left(\begin{array}{cc|cc} 0 & 1 & 0 & 2 \\ 2 & 2 & 1 & 17 \end{array}\right),$$

$$X^{(18)}\left(\begin{array}{cc|cc} 0 & 1 & 1 & 1 \\ 2 & 2 & 0 & 18 \end{array}\right), \quad X^{(19)}\left(\begin{array}{cc|cc} 0 & 1 & 0 & 2 \\ 3 & 1 & 0 & 18 \end{array}\right), \quad X^{(20)}\left(\begin{array}{cc|cc} 0 & 2 & 0 & 1 \\ 0 & 3 & 3 & 16 \end{array}\right), \quad X^{(21)}\left(\begin{array}{cc|cc} 0 & 2 & 1 & 0 \\ 0 & 3 & 2 & 17 \end{array}\right),$$

$$X^{(22)}\left(\begin{array}{cc|cc} 0 & 2 & 0 & 1 \\ 1 & 2 & 2 & 17 \end{array}\right), \quad X^{(23)}\left(\begin{array}{cc|cc} 0 & 2 & 1 & 0 \\ 1 & 2 & 1 & 18 \end{array}\right), \quad X^{(24)}\left(\begin{array}{cc|cc} 0 & 2 & 0 & 1 \\ 2 & 1 & 1 & 18 \end{array}\right), \quad X^{(25)}\left(\begin{array}{cc|cc} 0 & 2 & 1 & 0 \\ 2 & 1 & 0 & 19 \end{array}\right),$$

$$X^{(26)}\begin{pmatrix}0&2&0&1\\3&0&0&19\end{pmatrix},\quad X^{(27)}\begin{pmatrix}0&3&0&0\\0&2&3&17\end{pmatrix},\quad X^{(28)}\begin{pmatrix}0&3&0&0\\1&1&2&18\end{pmatrix},\quad X^{(29)}\begin{pmatrix}0&3&0&0\\2&0&1&19\end{pmatrix},$$

$$X^{(30)}\begin{pmatrix}1&0&0&2\\0&4&2&16\end{pmatrix},\quad X^{(31)}\begin{pmatrix}1&0&1&1\\0&4&1&17\end{pmatrix},\quad \ldots,\quad X^{(33)}\begin{pmatrix}1&0&0&2\\1&3&1&17\end{pmatrix},$$

$$X^{(34)}\begin{pmatrix}1&0&1&1\\1&3&0&18\end{pmatrix},\quad X^{(35)}\begin{pmatrix}1&0&0&2\\2&2&0&18\end{pmatrix},\quad X^{(36)}\begin{pmatrix}1&1&0&1\\0&3&2&17\end{pmatrix},\quad X^{(37)}\begin{pmatrix}1&1&1&0\\0&3&1&18\end{pmatrix},$$

$$X^{(38)}\begin{pmatrix}1&1&0&1\\1&2&1&18\end{pmatrix},\quad X^{(39)}\begin{pmatrix}1&1&1&0\\1&2&0&19\end{pmatrix},\quad X^{(40)}\begin{pmatrix}1&1&0&1\\2&1&0&19\end{pmatrix},\quad X^{(41)}\begin{pmatrix}1&2&0&0\\0&2&2&18\end{pmatrix},$$

$$X^{(42)}\begin{pmatrix}1&2&0&0\\1&1&1&19\end{pmatrix},\quad X^{(43)}\begin{pmatrix}1&2&0&0\\2&0&0&20\end{pmatrix},\quad X^{(44)}\begin{pmatrix}2&0&0&1\\0&3&1&18\end{pmatrix},\quad X^{(45)}\begin{pmatrix}2&0&1&0\\0&3&0&19\end{pmatrix},$$

$$X^{(46)}\begin{pmatrix}2&0&0&1\\1&2&0&19\end{pmatrix},\quad X^{(47)}\begin{pmatrix}2&1&0&0\\0&2&1&19\end{pmatrix},\quad X^{(48)}\begin{pmatrix}2&1&0&0\\1&1&0&20\end{pmatrix},\quad X^{(49)}\begin{pmatrix}3&0&0&0\\0&2&0&20\end{pmatrix}.$$

(In this particular example most of the arrays need to be written out because of the extremely small size of some of the border totals. The gain in time is greater as these increase, indeed had these rules been adopted in the case of the first example considered, only 11 of the 18 arrays there would have been written out.)

At the expense of not having a check by summing all the probabilities to make unity, a considerable further saving of time can be made. In order to determine $\wp_L$, only those values of $P_{X^{(t)}}$ such that $P_{X^{(t)}} \leqslant P_L$ need be found. Now if P_L is known and an approximate calculation shows that some $P_{X^{(t)}}$ is greater than P_L, or if both $P_{X^{(t')}} > P_L$ and $\phi(t', t) \geqslant 1$, then there is no need to calculate $P_{X^{(t)}}$.

Now
$$P_L = P_{X^{(6)}} = \frac{5!\,20!\,3!\,22!\,3!\,22!}{(25!)^2\,0!\,0!\,1!\,4!\,1!\,2!\,1!\,16!} = \frac{9\times 17\times 19}{23\times 23\times 100} = 0{\cdot}0551.$$

$$P_{X^{(5)}} = P_{X^{(6)}}\frac{(1\times 16)}{(2\times 3)} > P_{X^{(6)}},$$

$$P_{X^{(1)}} = P_{X^{(5)}}\frac{(1\times 15)}{(3\times 5)} = P_{X^{(5)}} > P_{X^{(6)}}.$$

It is thus easily seen that the values of t when $P_{X^{(t)}} > P_L$ are $t = 1, 2, 5, 11, 14$. Thus $\wp_L$ is the sum of $P_{X^{(t)}}$ for all values of t except 1, 2, 5, 11, 14. However, since in this example there are only 5 values of t not in $\wp_L$ as opposed to 44 in $\wp_L$ it is obviously simpler to calculate $1-\wp_L$, particularly since $P_{X^{(1)}}$ is already known.

We find $1-\wp_L = 0{\cdot}6353$ and so $\wp_L = 0{\cdot}3647$. There is thus no significant evidence of heterogeneity in the distribution.

(iv) An even more powerful though less frequently applicable way of eliminating calculation is the following, which could be used in the second of the examples considered here. Let there be N possible arrays $X^{(t)}$ and let the level of significance worked to be α.

Then if $P_L \leqslant \dfrac{\alpha}{N-1}$, the total probability $\wp_L$ must be less than α; and if $P_L \geqslant \alpha$, the total probability $\wp_L$ must be greater than α. In either case, the significance of the effect being tested is already determined and no more probabilities need be worked out, unless the value of $\wp_L$ is needed.

VII. The method described above is generally of use in cases where χ^2 would be used were not certain of the observed and expected numbers too small. However, it may also be used where χ^2 is wholly unsuitable, such a case being that described at the end of § VI (i). If this case alone had been considered, then the calculations, not lengthy in the whole example, would have been very much shortened as not all of the probabilities would have needed to be worked out.

The main use of the method is however in cases such as that described in the first part of § VI (i). It happens very frequently, particularly in cases where selection and training units are considered, that it is wished to compare their assessments of candidates when, under the conditions studied, the whole population contains so few members that χ^2 is inapplicable and still a test of significance is required. It was in order to study such cases that this technique was devised, and when the method is used as often as the frequency of occurrence of such cases demands it can be reduced to an almost mechanical simplicity by means of the devices used in § VI. The method is very easy to handle and all of the calculation is of a simple and straightforward character.

THE GEOMETRY OF ESTIMATION

By J. DURBIN and M. G. KENDALL

Introduction

1. The theory of statistical estimation is usually expounded by algebraic and analytical methods. Apart from the use of a multi-dimensional space by Neyman & Pearson (1928) in connexion with fitting theoretical distributions, by Hotelling (1930) and Fisher (1938) in a discussion of maximum likelihood estimators, and apart from a remark by Huzurbazar (1949) concerning the radius of curvature of a likelihood curve, there seem to have been few attempts to consider estimation as a geometrical problem. In the following paper we shall show how many of the familiar results in the theory of estimation can be represented simply and directly as properties of linear spaces, and that results so apparently different as Gauss's* theorem on least squares and Fisher's theorem concerning the minimal variance of maximum likelihood estimators in the limit are really aspects of one fundamental fact.

In the ultimate analysis geometrical 'proofs' in more than three dimensions are only restatements of analytical results in a special language; but they are nevertheless very useful, partly because of their elegance and partly because they carry a greater degree of conviction and understanding, to some minds at least, than the analytical approach. They also suggest generalizations and we shall, in fact, apply them to generalize Gauss's theorem to the case of the simultaneous estimation of several parameters.

Linear estimation and Gauss's theorem

2. Consider a set of n random variables $x_1, \ldots, x_n$ which are independent and have the same mean μ and variance σ^2. It is required to estimate the mean by a linear estimator

$$t = \sum_{j=1}^{n} \lambda_j x_j. \tag{1}$$

We impose the condition that this shall be unbiased, i.e. that

$$\Sigma \lambda_j = 1. \tag{2}$$

Throughout the whole of this paper we shall be concerned with one aspect or another of linear unbiased estimators of this type.

Consider now a Euclidean $[n]$ space with co-ordinates $\lambda_1, \ldots, \lambda_n$, which we call the *estimator* space. The hyperplane (2) corresponds to the range of values of λ giving unbiased estimators and any point P in it determines just one estimator. Now the variance of the estimator is $\sigma^2 \Sigma \lambda_j^2$ and hence is $\sigma^2 OP^2$ where O is the origin. It follows that this is a minimum when P is the foot of the perpendicular from O on to the hyperplane. Symmetry alone is enough to show that the values of the λ's are then all equal. Moreover, the geometrical approach shows at once that the value is a true minimum, not merely a stationary value, a point which the analytical or algebraical approach does not bring out immediately.

If P' is a point corresponding to any other estimator t' we have

$$\frac{\operatorname{var} t}{\operatorname{var} t'} = \frac{OP^2}{OP'^2} = \cos^2 \phi, \quad \text{say}, \tag{3}$$

* The theorem is often attributed to Markoff but in all its essentials is due to Gauss. See the historical note by R. L. Plackett, *Biometrika*, 1949, **36**, 458.

where ϕ is the angle POP'. But $\cos\phi$ is the correlation between t and t' and thus, if E is the efficiency of t' we have

$$\rho(t,t') = \sqrt{\frac{\text{var}\, t}{\text{var}\, t'}} = \sqrt{E}. \tag{4}$$

A general proof of this result is given by Cramér (1946).

We notice also from the geometry of the situation that if the λ's of t' are near the optimum values, namely those of t, the efficiency of t' is nearly unity. This is important for the practical utility of the result when we generalize it to the case of unequal variances below.

3. Consider now a second Euclidean $[n]$ space corresponding to the x's, which we call the *observation* space. The bilinear form $\Sigma\lambda_j x_j$ establishes a duality between this and the estimator space: for any fixed t, to a point in one there corresponds a hyperplane in the other. In particular, to the hyperplane (2) there corresponds the point $(t, t, \ldots, t)$, lying on the line $x_1 = x_2 = \ldots x_n$. Moreover, if a vector through the origin in one space is perpendicular to a hyperplane in that space, the corresponding hyperplane and vector in the other space are also orthogonal. If t is allowed to vary, then to a point in one space there will correspond a family of parallel hyperplanes in the other.

It follows that the unbiased estimator of minimum variance in the observation space is given by the family of hyperplanes orthogonal to the line $x_1 = x_2 = \ldots x_n$. Thus if Q is the point $(x_1 \ldots x_n)$ the optimum estimate is given by dropping a perpendicular on to this line, meeting it at L, say. Without loss of generality we may take the origin of the observation space, say X, to be the mean, and the estimate is then XL. This procedure is equivalent to finding L such that QL^2 is a minimum. In other language we choose the estimate so as to minimize $\sum_{j=1}^{n}(x_j - t)^2$, i.e. take the least-squares solution for t.

This is Gauss's theorem in its simplest form. We minimize a sum of squares and emerge with an optimum estimator, namely one which minimizes another sum of squares (the variance). From the geometrical viewpoint this arises simply from the duality between two spaces established by the bilinear form of the type of estimator we are considering.

4. We can now proceed to generalize the simple form. Suppose the x's come from different populations with variances σ_i^2 which are known except for some constant factor k^2 (i.e. we know each $k^2\sigma_i^2$) and with means $\mu_i = \beta a_i$ where the a_i's are known and our object is to estimate β. (This is the model for a regression through the origin.) We first of all transform to new variables

$$\xi_i = \frac{x_i}{k\sigma_i}. \tag{5}$$

The ξ's now have equal variances and means $\beta a_i/k\sigma_i$ say βc_i where the c_i's are known. Geometrically this means a contraction of the variate scales. Taking now

$$t = \Sigma\lambda_i\xi_i \tag{6}$$

as an estimator of β we see from the foregoing argument that the optimum unbiased estimator is given in the estimator space by the vector through the origin perpendicular to the hyperplane

$$\Sigma\lambda_i c_i = 1, \tag{7}$$

that is to say, by choosing the λ's to be proportional to the c's. The optimum is thus given by

$$\begin{aligned} b &= \Sigma c_i\xi_i / \Sigma c_i^2 \\ &= \Sigma\frac{a_i x_i}{k^2\sigma_i^2} \Big/ \Sigma\frac{a_i^2}{k^2\sigma_i^2}. \end{aligned} \tag{8}$$

In the observation space the estimator is given by the family of planes orthogonal to the line

$$\frac{\xi_1}{c_1} = \ldots = \frac{\xi_n}{c_n}, \tag{9}$$

that is to say, by minimizing the sum

$$\sum_{i=1}^{n} (\xi_i - \beta c_i)^2 \equiv \sum_{i=1}^{n} \left(\frac{x_i - \beta a_i}{k\sigma_i}\right)^2 \tag{10}$$

which leads to the same formula (8).

5. Before proceeding to further generalizations we introduce an alternative geometrical representation for the case just considered, except that we assume the variates already standardized by division by $k\sigma_i$. This involves no loss of generality.

The observation vector $\mathbf{x}$ and the estimator vector $\boldsymbol{\lambda}$ can be represented by the vectors OX and OL in a single Euclidean $[n]$ space. Let T be the foot of the perpendicular from X to OL. Then $t = OT/OL$ (see Fig. 1).

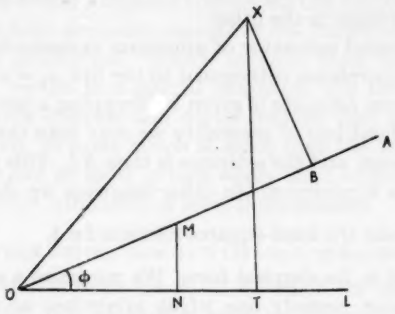


Fig. 1

Let OM and OA represent the vectors $\boldsymbol{\mu}$ and $\mathbf{a}$. M must lie on OA since $\mu_i = \beta a_i$. The least-squares estimator is the value of b which minimizes the sum of squares $\Sigma(x_i - ba_i)^2$. Let OB represent $b\mathbf{a}$. Then $\Sigma(x_i - ba_i)^2 = XB^2$. Thus B is the projection of X on OA. Also $b = OB/OA$. MX represents the vector $\boldsymbol{\epsilon} = \mathbf{x} - \beta\mathbf{a}$. $\epsilon_1, \ldots, \epsilon_n$ are uncorrelated random variables with zero means and constant variance σ^2. Any orthogonal transformation of them preserves distances and inner products, and so gives new variables that are uncorrelated with zero means and the same variance. Choose such a transformation in which one of the new axes coincides with OA. Then MB is the corresponding new variable. Thus MB has zero mean and variance σ^2, i.e. OB has mean OM and variance σ^2, i.e. b has mean β and variance σ^2/OA^2.

Let N be the projection of M on OL. Make a new orthogonal transformation such that one axis coincides with OL. The corresponding new variable is NT which, as before, has zero mean and variance σ^2. Thus OT has mean ON and variance σ^2, i.e. tOL has mean $\beta\, OA \cos\phi$ and variance σ^2 where ϕ is the angle MON. Thus if t is an unbiased estimator of β we must have $OA \cos\phi = OL$ in which case

$$\operatorname{var} t = \frac{\sigma^2}{OA^2 \cos^2\phi} = \frac{\operatorname{var} b}{\cos^2\phi}.$$

This is the result obtained by the earlier method. As before $\cos\phi = \operatorname{corr}(b, t)$.

6. This approach gives another explanation of why the least-squares estimator has least variance. Any linear estimator can be obtained as a multiple, say k, of the projection of the sample vector upon a fixed vector. The mean of the estimator is simply k times the projection of the centroid of the distribution upon the fixed vector. For comparative purposes we can confine ourselves to fixed vectors of given length. As the fixed vector departs from the vector passing through the centroid of the distribution the mean of the sample projection upon it diminishes. Thus if the estimator is to be unbiased the value of k must be increased. This has the effect of increasing the variance of the estimator. It can be seen that k is least when the fixed vector passes through the centroid of the distribution. The projection upon this vector gives the least-squares estimator. Thus the least-squares estimator is also the best linear unbiased estimator.

GAUSS'S THEOREM ON THE ESTIMATION OF A LINEAR FUNCTION OF SEVERAL PARAMETERS

7. If the variates have different variances we assume them standardized in the manner of the preceding sections. A further generalization of Gauss's theorem then concerns the estimation of some linear function of parameters β which in turn are linearly connected with the means of the variates. The set-up is the same as in §4, except that $\mu_i = \sum_{j=1}^{k} \beta_j a_{ji} \ (i = 1, 2, \ldots, n)$, or in matrix notation $\boldsymbol{\mu} = \mathbf{A}\boldsymbol{\beta}$ where $\mathbf{A}$ is a known matrix. The problem is to estimate the form $\delta = \sum_{1}^{k} d_j \beta_j$, i.e. $\mathbf{d}'\boldsymbol{\beta}$ where the d's are known constants, the β's being unknown. The theorem states that the estimator obtained by substituting in the expression for δ the least-squares estimators of $\beta_1, \beta_2, \ldots, \beta_k$ is the best linear unbiased estimator of δ.

8. *Duality principle.* By the usual methods we find that the least-squares estimator of δ is given by

$$\mathbf{d} = \mathbf{d}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{x}. \tag{11}$$

This is a bilinear form which represents an $[n-1]$ hyperplane in the observation space which is perpendicular to the vector $\mathbf{d}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$.

Suppose $t = \sum_{1}^{n} \lambda_i x_i$ is any estimator of δ. If this is unbiased we must have

$$E(t) = \boldsymbol{\lambda}'\mathbf{A}\boldsymbol{\beta} = \boldsymbol{\delta} = \mathbf{d}'\boldsymbol{\beta}$$

identically, if t is to be unbiased for all values of the β's. Thus we must have

$$\boldsymbol{\lambda}'\mathbf{A} = \mathbf{d}'. \tag{12}$$

Now

$$\operatorname{var} t = \sigma^2 \Sigma \lambda_i^2$$

is a minimum subject to (12) if the vector $\boldsymbol{\lambda}$ in the estimator space is perpendicular to the $[n-k]$ hyperplane specified by (12). The condition for this is that $\boldsymbol{\lambda}' = \mathbf{d}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$.

Once again the estimators are the same since the process of minimizing the sum of squares in the observation space is the dual of the process of minimizing the variance in the estimator space. Both processes involve minimizing the distance from a point to a subspace and therefore give rise to an orthogonal partition of the space.

As before if t is any unbiased linear estimator

$$\operatorname{var} t = \frac{\operatorname{var} b}{\cos^2 \phi},$$

where ϕ is the angle between the two vectors of coefficients in the estimator space. Also $\cos\phi = \text{corr}\,(b, t)$. Equation (4) is therefore generally true.

9. *Alternative representation.* For simplicity we shall consider the case $k = 2$, though the arguments are essentially quite general. Let OX, OM, OA_1 and OA_2 represent the vectors $\mathbf{x}$, $\boldsymbol{\mu}$, $\mathbf{a}_1$ and $\mathbf{a}_2$, and along OA_1 and OA_2 take $OM_1 = \beta_1 OA_1$ and $OM_2 = \beta_2 OA_2$ respectively. Let B be the projection of X on the [2] space spanned by OA_1 and OA_2 with OB_1BB_2 a quadrilateral as shown in Fig. 2. Then $b_1 = \dfrac{OB_1}{OA_1}$ and $b_2 = \dfrac{OB_2}{OA_2}$ are the least squares estimators of β_1 and β_2.

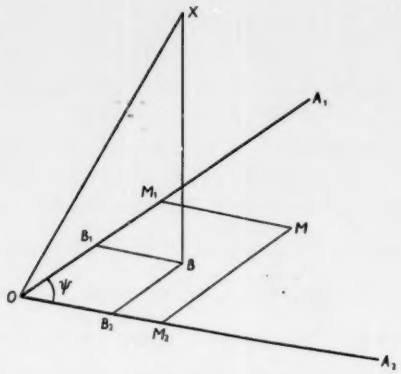


Fig. 2

As before, MX represents the vector $\boldsymbol{\epsilon} = \mathbf{x} - \boldsymbol{\mu}$. By making a linear transformation of the ϵ's such that two of the new axes coincide with A_1 and A_2, while the remaining axes are orthogonal to these, we see that the means of OB_1 and OB_2 are OM_1 and OM_2 respectively. Thus b_1 and b_2 are unbiased. Consequently, the least-squares estimator $d = d_1b_1 + d_2b_2$ is an unbiased estimator of $\delta = d_1\beta_1 + d_2\beta_2$.

Suppose $t = \Sigma\lambda_i x_i$ is any unbiased linear estimator of δ. As before $\text{var}\,t$ is least when the vector $\boldsymbol{\lambda}$ passes through the centroid M. The length

$$OM = \sqrt{(\beta_1^2 OA_1^2 + \beta_2^2 OA_2^2 + 2\beta_1\beta_2 OA_1 OA_2 \cos\psi)},$$

where ψ is the angle A_1OA_2. Thus the estimator of δ with least variance is

$$\frac{\delta}{OM}\sum_1^n (\beta_1 a_{1i} + \beta_2 a_{2i})\, x_i.$$

This estimator is of no use in practice, however, since it requires a knowledge of β_1 and β_2 which are, of course, unknown. What we require is the best among those unbiased estimators whose coefficients do not depend upon β_1 and β_2.

As before, let OL represent the vector $\boldsymbol{\lambda}$ corresponding to any estimator $t = \Sigma\lambda_i x_i$ and let OT, ON be the projections of X, M on OL. Then OT has mean ON and variance σ^2. Now $ON = \beta_1 OA_1 \cos\theta_1 + \beta_2 OA_2 \cos\theta_2$, where θ_1 is the angle LOA_1 and θ_2 is the angle LOA_2. We require as an estimator of δ some multiple, say k, of OT. If this is to be unbiased we must have $k(\beta_1 OA_1 \cos\theta_1 + \beta_2 OA_2 \cos\theta_2) = d_1\beta_1 + d_2\beta_2$. If, in addition, t is to be independent

of β_1 and β_2 we must have $\frac{OA_1 \cos\theta_1}{OA_2 \cos\theta_2} = \frac{d_1}{d_2}$ which is a constant. Thus OL is restricted to move in an $n-1$ space orthogonal to the [2] space spanned by OA_1 and OA_2.

Of the lines through the origin in this space the one giving the minimum variance estimator is the line that passes closest to M. Clearly this is the intersection of the space with OA_1A_2. There is only one such intersection and this is the line in the space OA_1A_2 that gives an unbiased estimator, that is, it is the line giving the least-squares estimator since this is known to be unbiased. Consequently, the least-squares estimator is the best linear unbiased estimator.

As before, if $t = \Sigma\lambda_i x_i$ is any other estimator $\operatorname{var} t = (\operatorname{var} d/\cos^2\phi)$, where ϕ is the angle between the two vectors of coefficients. Thus ϕ is the angle between the vector $\boldsymbol{\lambda}$ and the space OA_1A_2, that is, $\cos\phi$ is the multiple correlation between $\boldsymbol{\lambda}$ and $\mathbf{a}_1$ and $\mathbf{a}_2$.

In § 15 we present a generalization of Gauss's theorem to the case when several parameters are simultaneously under estimate.

10. *Correlated variates.* The case when the x's are not independent has been dealt with very elegantly from the algebraical stand-point by Aitken (1935). Geometrically, the duality given by the orthogonal properties in the independent case is replaced by a more complex position depending on angles. For instance, in the simple case of § 2, the optimum point in the hyperplane is given by the point of contact P with the ellipsoid

$$\Sigma\rho_{ij}\lambda_i\lambda_j = \text{constant}, \tag{13}$$

where ρ_{ij} is the covariance of x_i and x_j. To this point there corresponds a family of parallel planes in the observation space making an angle with the line $x_1 = x_2 = \ldots x_n$ equal to the angle between OP and the hyperplane in the estimator space.

The use of such results depends, however, on a knowledge of the covariance matrix ρ_{ij} (except for a multiplicative constant) and if the matrix is in fact known it is simpler to 'orthogonalize' the problem by transforming to new variates which are independent. This amounts to linear deformations of the spaces; and as such transformations leave linear subspaces invariant the general results for the case of independence are linearly transformable to the case of dependence. In the latter case, instead of minimizing a sum of squares we minimize a quadratic form in order to give the optimum unbiased estimator.

11. *Estimates of the variance.* In the simple case of § 2 the variance of the optimum estimator t is given immediately as OP^2 namely σ^2/n. In the more general case of § 7 it is $\sigma^2\Sigma\lambda_i^2$, where $\boldsymbol{\lambda}' = \mathbf{d}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. What is more usually required is an estimator of these quantities and the second part of Gauss's theorem, in its extended form, states that an unbiased estimator is given by

$$\frac{S}{n-k}\Sigma\lambda_i^2, \tag{14}$$

where S is the minimum value of the sum of squares used in arriving at the λ's. This will follow if we show that $S/(n-k)$ is an unbiased estimator of σ^2, which is readily seen to be so from the consideration that S is the square of the perpendicular from a point in the observation space on to an $[n-k]$ space.

MAXIMUM-LIKELIHOOD ESTIMATORS

12. We now consider a third geometrical representation. We regard the parent frequency distribution as divided into m groups with relative frequencies, say, $a_1, \ldots, a_m$. For a sample of n the observed proportional frequencies in each group, say, $n_1, \ldots, n_m$, will fluctuate about the a's subject to the condition

$$\sum_{j=1}^{m} n_j = 1. \tag{15}$$

We consider a Euclidean $[m]$ space in which the co-ordinates are $n_1, \ldots, n_m$. The 'parent' point, say P, has co-ordinates $a_1, \ldots, a_m$. The sample points will cluster round P and as n becomes large they become a globular cluster for which the variables

$$x_i = \frac{n_i - a_i}{\sqrt{a_i}} \tag{16}$$

are normally distributed in the $[m-1]$ space α defined by $\Sigma x_i \sqrt{a_i} = 0$ with unit variance and zero mean. A representation of this kind was considered by Neyman & Pearson (1928), and later by Hotelling (1930) and Fisher (1938).

Now for any value of θ, a parameter of the parent, there will be a corresponding point Q and the locus of such point provides a curve which, following Neyman & Pearson we call the Population Locus. P itself lies on this curve. The principle of maximum likelihood requires that if S is the sample point we estimate θ by dropping a perpendicular in α from S on to this curve, meeting it at T, say. The estimate of θ is the value corresponding to the point T. To put it another way, the estimator t of θ is defined by the family of hyperplanes in α normal to the population locus through the sample points.

13. In the limit the sample points are clustered very closely round P and the population locus in the neighbourhood of P is linear. Also the variation of θ as a function of distance along the curve tends to linearity. The maximum-likelihood estimator is therefore given by hyperplanes normal to a straight line. On transforming to variates which are independent and have equal variance, it follows from Gauss's theorem as proved in § 2 that the maximum-likelihood estimator has minimum variance among all linear unbiased estimators, that is to say, among all estimators which are based on linear functions of the relative frequencies and are consistent. This is Fisher's theorem.

14. Consider now the case when several parameters are simultaneously under estimate. Instead of a population curve we shall have a population surface of k dimensions, where k is the number of parameters. The maximum likelihood estimator is given by dropping perpendiculars on to this surface, which again for large samples is flat in the effective range of variation. To anticipate the argument in the following section, this is equivalent to minimizing the volume of an ellipsoid or, in the dual form, to minimizing the volume of a hyperpyramid. Both these are equivalent in turn to minimizing the determinant of the covariance matrix. Hence we get Geary's theorem (1942): the maximum likelihood estimators in large samples minimize the generalized variance

$$V = |\operatorname{cov} t_i, t_j|, \tag{17}$$

where t_i is the estimator of θ_i and $\operatorname{cov}(t_i, t_i) = \operatorname{var} t_i$.

SIMULTANEOUS LINEAR ESTIMATION

15. To illustrate the power of the geometrical method still further we shall generalize Gauss's theorem to the case when several parameters (or linear functions of them) are simultaneously under estimate. Suppose that we have the same model as in §9, i.e. $\mathbf{y}$ is a vector of uncorrelated standardized variates with means given by $E(\mathbf{y}) = \mathbf{A}\boldsymbol{\beta}$. We require to estimate the p linear functions of the β's given by $\boldsymbol{\delta} = \mathbf{D}\boldsymbol{\beta}$, where $\boldsymbol{\delta} = \{\delta_1, \ldots, \delta_p\}$ and $\mathbf{D}$ is a known $p \times k$ matrix of rank $\leqslant k$. Our generalized form of Gauss's theorem states that the linear unbiased estimators which have least generalized variance are obtained by substituting the least-squares estimators of $\beta_1, \beta_2, \ldots, \beta_k$ in the expression for $\boldsymbol{\delta}$.

An algebraic proof of this proposition has been given for the case $\mathbf{D} = \mathbf{I}$ by Aitken (1948). The following derivation may, however, be found easier to grasp since Aitken's treatment requires rather sophisticated algebra.

We shall first prove the result for the case $\mathbf{D} = \mathbf{I}$, the unit matrix, that is, we shall show that the least-squares estimator $\mathbf{b} = \{b_1, \ldots, b_k\}$ of $\boldsymbol{\beta}$ has minimum generalized variance in the class of unbiased estimators of $\boldsymbol{\beta}$.

First we note that the generalized variance of a set of variates $\mathbf{z}$ with covariance matrix $\mathbf{V}$ is proportional to the square of the volume of the ellipsoid $\mathbf{z}'\mathbf{V}\mathbf{z} =$ constant. This ellipsoid is sometimes called the ellipsoid of concentration; it may be constructed for the least-squares estimators in the following way.

Using the same geometrical representation as in §9, except that the number k of columns of $\mathbf{A}$ is unrestricted, let α be the space spanned by $OA_1, \ldots, OA_k$. Then $b_1, \ldots, b_k$ are constant multiples of the co-ordinates of B, the orthogonal projection of X on α, referred to the oblique axes $OA_1, \ldots, OA_k$. For simplicity suppose that each of these multiples is unity. The orthogonal projection upon α of an $[n]$ sphere C, whose centre is the centroid M of the distribution of the y's, is a $[k]$ sphere, S say. The sphere S referred to the axes $OA_1, \ldots, OA_k$ gives the ellipsoid of concentration of $b_1, \ldots, b_k$.

Let $\mathbf{t} = \boldsymbol{\Lambda}\mathbf{y}$ be any linear estimator of $\boldsymbol{\beta}$. This matrix equation determines an $[n-k]$ space, all points of which give the same value of $\mathbf{t}$ and which intersects α in a point T'. The co-ordinates of T' relative to the axes $OA_1, \ldots, OA_k$ evidently depend linearly upon the elements $t_1, \ldots, t_k$ of $\mathbf{t}$. As $\mathbf{t}$ varies these co-ordinates have means equal to the co-ordinates of the centroid of the distribution of the y's, that is they have means $\beta_1, \ldots, \beta_k$. Since $\mathbf{t}$ has been defined to have mean $\boldsymbol{\beta}$, $t_1, \ldots, t_k$ must therefore be the co-ordinates of T'.

Every point Y in the space $\mathbf{t} = \boldsymbol{\Lambda}\mathbf{y}$ gives the same point T'. As Y moves over the sphere C, T' traces out a $[k]$ ellipsoid, say E, in α. E referred to the axes $OA_1, \ldots, OA_k$ gives the ellipsoid of concentration of $\mathbf{t}$.

It follows that the sphere S must either coincide with E or must lie inside it. Consequently the ellipsoid of concentration of $\mathbf{b}$ must either coincide with that of $\mathbf{t}$ or lie inside it. We conclude that the generalized variance of $\mathbf{t}$ cannot be less than that of $\mathbf{b}$.

The general case of estimation of the linear forms $\boldsymbol{\delta} = \mathbf{D}\boldsymbol{\beta}$ can be dealt with in a similar way though the argument is more complicated. First it must be shown that the least-squares estimates $\mathbf{d} = \{d_1, \ldots, d_p\}$ can be obtained as the co-ordinates relative to certain oblique axes of the projection of the sample point on the $[p]$ space spanned by the axes. Any other estimator $\mathbf{t} = \boldsymbol{\Lambda}\mathbf{y}$ represents an $[n-p]$ space which meets the $[p]$ space in a point T'. By considering the conditions for unbiasedness it can be shown that $t_1, \ldots, t_p$ are the oblique co-ordinates of T'. As before, the ellipsoid of concentration of $\mathbf{d}$ coincides

with that of **t** or lies inside it. Hence the generalized variance of **t** cannot be less than that of **d**.

From the statement of the theorem we also have the notable result that minimization of the covariance determinant is equivalent to minimization of each variance separately, subject to the condition of unbiasedness, and also to the minimization of any principal minor of the determinant.

REFERENCES

AITKEN, A. C. (1935). *Proc. Roy. Soc. Edinb.* **55**, 42.
AITKEN, A. C. (1948). *Proc. Roy. Soc. Edinb.* **62**, 369.
CRAMÉR, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press.
FISHER, R. A. (1938). *Theory of Statistical Estimation*, University of Calcutta.
GEARY, R. C. (1942). *J.R. Statist. Soc.* **105**, 213.
HOTELLING, H. (1930). *Trans. Amer. Math. Soc.* **32**, 847.
HUZURBAZAR, V. S. (1949). *Biometrika*, **36**, 71.
NEYMAN, J. & PEARSON, E. S. (1928). *Biometrika*, A **20**, 175, 263.

TESTING FOR SERIAL CORRELATION IN LEAST SQUARES REGRESSION. II

By J. DURBIN,* *London School of Economics*
AND
G. S. WATSON, *Department of Applied Economics, University of Cambridge*

1. Introduction

In an earlier paper (Durbin & Watson, 1950) the authors investigated the problem of testing the error terms of a regression model for serial correlation. Test criteria were put forward, their moments calculated, and bounds to their distribution functions were obtained. In the present paper these bounds are tabulated and their use in practice is described. For cases in which the bounds do not settle the question of significance an approximate method is suggested. Expressions are given for the mean and variance of a test statistic for one- and two-way classifications and polynomial trends, leading to approximate tests for these cases. The procedures described should be capable of application by the practical worker without reference to the earlier paper (hereinafter referred to as Part I).

It should be emphasized that the tests described in this paper apply only to regression models in which the independent variables can be regarded as 'fixed variables'. They do not, therefore, apply to autoregressive schemes and similar models in which lagged values of the dependent variable occur as independent variables.

2. The bounds test

Throughout the paper the procedures suggested will be illustrated by numerical examples. We begin by considering some data from a demand analysis study.

Example 1. *Annual consumption of spirits from* 1870 *to* 1938. The data (given in Table 1) were compiled by A. R. Prest, to whose paper (1949) reference should be made for details of the source material. As is common in econometric work the original observations were transformed by taking logarithms:

y = log consumption of spirits per head;

x_1 = log real income per head;

x_2 = log relative price of spirits (i.e. price of spirits deflated by a cost-of-living index).

We suppose that the observations satisfy the regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \tag{1}$$

where β_0 is a constant, β_1 is the income elasticity, β_2 is the price elasticity, and ϵ is a random error with zero mean and constant variance.

To test the errors for serial correlation the following sums of squares and products are required:

$\Sigma(y-\bar{y})^2$	= 5·000123	$\Sigma(y-\bar{y})(x_2-\bar{x}_2)$	= −3·763579	$\Sigma(\Delta x_2)^2$	= 0·083559
$\Sigma(x_1-\bar{x}_1)^2$	= 0·632006	$\Sigma(x_1-\bar{x}_1)(x_2-\bar{x}_2)$	= 1·014984	$\Sigma\Delta y\Delta x_1$	= 0·014685
$\Sigma(x_2-\bar{x}_2)^2$	= 2·966354	$\Sigma(\Delta y)^2$	= 0·112592	$\Sigma\Delta y\Delta x_2$	= −0·076399
$\Sigma(y-\bar{y})(x_1-\bar{x}_1)$	= −1·321973	$\Sigma(\Delta x_1)^2$	= 0·023539	$\Sigma\Delta x_1\Delta x_2$	= 0·000527

* Junior research worker at the Department of Applied Economics, Cambridge, when this work was begun.

$\Sigma(\Delta y)^2$ stands for the sum of squares of the first differences of the y's, and $\Sigma \Delta y \Delta x$ stands for the sum of products of the first differences of the y's times the corresponding first differences of the x's, etc. Thus the first term in $\Sigma(\Delta y)^2$ is $(1{\cdot}9794 - 1{\cdot}9565)^2 = 0{\cdot}00052441$, and the first term in $\Sigma \Delta y \Delta x_1$ is $(1{\cdot}9794 - 1{\cdot}9565)(1{\cdot}7766 - 1{\cdot}7669) = 0{\cdot}00022213$. Since there are 69 observations there are 69 terms in each of the first six summations, and 68 in each of the second six summations.

Table 1. *Annual consumption of spirits from* 1870 *to* 1938

Year	Consumption y	Income x_1	Price x_2	Year	Consumption y	Income x_1	Price x_2
1870	1·9565	1·7669	1·9176	1905	1·9139	1·9924	1·9952
1871	1·9794	1·7766	1·9059	1906	1·9091	2·0117	1·9905
1872	2·0120	1·7764	1·8798	1907	1·9139	2·0204	1·9813
1873	2·0449	1·7942	1·8727	1908	1·8886	2·0018	1·9905
1874	2·0561	1·8156	1·8984	1909	1·7945	2·0038	1·9859
1875	2·0678	1·8083	1·9137	1910	1·7644	2·0099	2·0518
1876	2·0561	1·8083	1·9176	1911	1·7817	2·0174	2·0474
1877	2·0428	1·8067	1·9176	1912	1·7784	2·0279	2·0341
1878	2·0290	1·8166	1·9420	1913	1·7945	2·0359	2·0255
1879	1·9980	1·8041	1·9547	1914	1·7888	2·0216	2·0341
1880	1·9884	1·8053	1·9379	1915	1·8751	1·9896	1·9445
1881	1·9835	1·8242	1·9462	1916	1·7853	1·9843	1·9939
1882	1·9773	1·8395	1·9504	1917	1·6075	1·9764	2·2082
1883	1·9748	1·8464	1·9504	1918	1·5185	1·9965	2·2700
1884	1·9629	1·8492	1·9723	1919	1·6513	2·0652	2·2430
1885	1·9396	1·8668	2·0000	1920	1·6247	2·0369	2·2567
1886	1·9309	1·8783	2·0097	1921	1·5391	1·9723	2·2988
1887	1·9271	1·8914	2·0146	1922	1·4922	1·9797	2·3723
1888	1·9239	1·9166	2·0146	1923	1·4606	2·0136	2·4105
1889	1·9414	1·9363	2·0097	1924	1·4551	2·0165	2·4081
1890	1·9685	1·9548	2·0097	1925	1·4425	2·0213	2·4081
1891	1·9727	1·9453	2·0097	1926	1·4023	2·0206	2·4367
1892	1·9736	1·9292	2·0048	1927	1·3991	2·0563	2·4284
1893	1·9499	1·9209	2·0097	1928	1·3798	2·0579	2·4310
1894	1·9432	1·9510	2·0296	1929	1·3782	2·0649	2·4363
1895	1·9569	1·9776	2·0399	1930	1·3366	2·0582	2·4552
1896	1·9647	1·9814	2·0399	1931	1·3026	2·0517	2·4838
1897	1·9710	1·9819	2·0296	1932	1·2592	2·0491	2·4958
1898	1·9719	1·9828	2·0146	1933	1·2635	2·0766	2·5048
1899	1·9956	2·0076	2·0245	1934	1·2549	2·0890	2·5017
1900	2·0000	2·0000	2·0000	1935	1·2527	2·1059	2·4958
1901	1·9904	1·9939	2·0048	1936	1·2763	2·1205	2·4838
1902	1·9752	1·9933	2·0048	1937	1·2906	2·1205	2·4636
1903	1·9494	1·9797	2·0000	1938	1·2721	1·1182	2·4580
1904	1·9332	1·9772	1·9952				

The regression coefficients are calculated by inverting the matrix

$$\begin{bmatrix} \Sigma(x_1-\bar{x}_1)^2 & \Sigma(x_1-\bar{x}_1)(x_2-\bar{x}_2)^2 \\ \Sigma(x_1-\bar{x}_1)(x_2-\bar{x}_2) & \Sigma(x_2-\bar{x}_2)^2 \end{bmatrix},$$

giving for the estimates of β_1 and β_2

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 2{\cdot}966354 & -1{\cdot}014984 \\ -1{\cdot}014984 & 0{\cdot}632006 \end{bmatrix} \begin{bmatrix} -1{\cdot}321973 \\ -3{\cdot}763579 \end{bmatrix},$$

i.e.

$$b_1 = -0{\cdot}120142, \quad b_2 = -1{\cdot}227647.$$

Let z denote the residual from regression, i.e.

$$z = y - \bar{y} - b_1(x_1 - \bar{x}_1) - b_2(x_2 - \bar{x}_2).$$

Then
$$\Sigma z^2 = \Sigma(y-\bar{y})^2 - b_1\Sigma(y-\bar{y})(x_1-\bar{x}_1) - b_2\Sigma(y-\bar{y})(x_2-\bar{x}_2)$$
$$= 0{\cdot}22095.$$

The statistic to be used for testing for serial correlation is

$$d = \frac{\Sigma(\Delta z)^2}{\Sigma z^2}. \tag{2}$$

The reasons for choosing this statistic have been given in Part I and need not be discussed here. Now

$$\Delta z = \Delta y - b_1\Delta x_1 - b_2\Delta x_2,$$

so that

$$\Sigma(\Delta z)^2 = \Sigma(\Delta y)^2 + b_1^2\Sigma(\Delta x_1)^2 + b_2^2\Sigma(\Delta x_2)^2 - 2b_1\Sigma\Delta y\Delta x_1 - 2b_2\Sigma\Delta y\Delta x_2 + 2b_1b_2\Sigma\Delta x_1\Delta x_2$$
$$= 0{\cdot}054967.$$

Substituting in (2) we have $d = 0{\cdot}2488$.

We must now decide what departures from the null hypothesis of serial independence of the errors ϵ need be considered. Experience with econometric data such as the present indicates a test against the existence of positive serial correlation. If the errors were positively serially correlated, d would tend to be relatively small, while if the errors were negatively serially correlated d would tend to be large. We therefore require a critical value of d, say d^*, such that if the observed value of d is less than d^* we may infer that positive serial correlation is established at the significance level concerned.

It was shown in Part I that exact critical values of this kind cannot be obtained. However, it is possible to calculate upper and lower bounds to the critical values. These are denoted by d_U and d_L. If the observed d is less than d_L we conclude that the value is significant, while if the observed d is greater than d_U we conclude that the value is not significant at the significance level concerned. If d lies between d_L and d_U the test is inconclusive.

Significance points of d_L and d_U are tabulated for various levels in Tables 4, 5 and 6. In addition, a diagram is given to facilitate the test procedure in the most usual case of a test against positive serial correlation at the 5 % level (Fig. 1). k' is the number of independent variables.

In the present example $n = 69$ and $k' = 2$, so that at the 5 % level $d_L = 1{\cdot}54$ approximately. The observed value 0·25 is less than this and therefore indicates significant positive serial correlation at the 5 % level. In fact, the observed value is also significant at the 1 % level.

The procedure for other values of k' is exactly similar. In all cases the value of d given by (2) is calculated, the z's being the residuals from regression, and the appropriate table is consulted.

Tests against negative serial correlation and two-sided tests

Tests against negative serial correlation may sometimes be required. For instance, it is a common practice in econometric work to analyse the first differences of the observations rather than the observations themselves, on the ground that the serial correlation of the transformed errors is likely to be less than that of the original errors. We may wish to ensure that the transformation has not overcorrected, thus introducing negative serial correlation

into the transformed errors. To make a test against negative serial correlation, d is calculated as above and subtracted from 4. The quantity $4-d$ may now be treated as though it were the value of a d-statistic to be tested for positive serial correlation. Thus if $4-d$ is less than d_L, there is significant evidence of negative serial correlation, and if $4-d$ is greater than d_U, there is not significant evidence; otherwise the test is inconclusive.

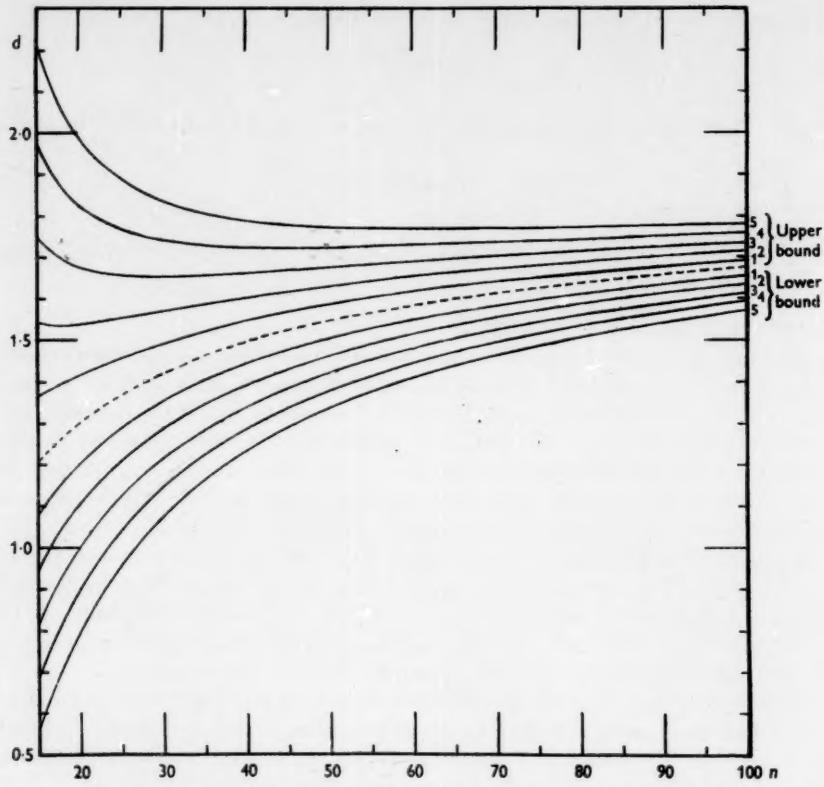


Fig. 1. Graphs of 5% values of d_L and d_U against n for $k'=1, 2, 3, 4, 5$.

When there is no prior knowledge of the sign of the serial correlation, two-sided tests may be made by combining single-tail tests. Using only equal tails, d will be significant at level α if either d is less than d_L or $4-d$ is less than d_L, non-significant if d lies between d_U and $4-d_U$ and inconclusive otherwise; the level α may be 2, 5 and 10 %. Thus, by using the 5 % values of d_L and d_U from Table 4, a two-sided test at the 10 % level is obtained.

3. Regression through the origin

The procedures described so far apply only to cases in which means have been fitted in the regressions, i.e. the fitted regression equations take the form

$$y-\bar{y} = b_1(x_1-\bar{x}_1)+\ldots+b_{k'}(x_{k'}-\bar{x}_{k'}). \tag{3}$$

These are the most common cases in practice. However, we occasionally require a fitted regression through the origin of the form

$$y = B_1 x_1 + \ldots + B_{k'} x_{k'}. \tag{4}$$

Tables 4, 5 and 6 do not apply directly to the residuals from regressions of this type. To test for serial correlation in such cases an equation of the form (3) must first be fitted. The residuals from the resulting regression may then be tested by the procedures of §§ 2 and 4. This gives a perfectly valid test even though (4) might be the more appropriate regression to fit for other purposes.

In order to avoid inverting more than one matrix the following method, due to Cochran (1938), should be used. First of all, the regression coefficients of equation (4) are determined. In this operation the inverse of the matrix of squares and cross-products of $x_1, \ldots, x_{k'}$ will be calculated; denote this matrix by $\mathbf{C} = \{c_{ij}\}$ and the means of the variables y, $x_1, \ldots, x_{k'}$ by $\bar{y}, \bar{x}_1, \ldots, \bar{x}_{k'}$. Then

$$\begin{bmatrix} b_1 \\ \vdots \\ b_{k'} \end{bmatrix} = \begin{bmatrix} B_1 \\ \vdots \\ B_{k'} \end{bmatrix} - nB_0 \mathbf{C} \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_{k'} \end{bmatrix},$$

where

$$B_0 = \frac{\sum_{i=1}^{k'} B_i \bar{x}_i - \bar{y}}{\sum_{i,j=1}^{k'} c_{ij} \bar{x}_i \bar{x}_j - 1}.$$

4. Approximate procedure when the bounds test is inconclusive

No satisfactory procedure of general application seems to be available for cases in which the bounds test is inconclusive. However, an approximate test can be made, and this should be sufficiently accurate if the number of degrees of freedom is large enough, say greater than 40. For smaller numbers this test can only be regarded as giving a rough indication.

The method used is to transform d so that its range of variation is approximately from 0 to 1 and to fit a Beta distribution with the same mean and variance. The mean and variance of d vary according to the values of the independent variables, so the first step is to calculate them for the particular case concerned. The method of calculation will be illustrated by means of the data of Example 1, although in practice an approximate test would not be required for this case since the bounds test has given a definite answer.

The description of the computing procedure is greatly facilitated by the introduction of matrix notation. Thus the set $\{y_1, y_2, \ldots, y_n\}$ of observations of the independent variable is denoted by the column vector $\mathbf{y}$. In the same way the set

$$\begin{bmatrix} x_{11} & x_{21} & \ldots & x_{k'1} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \ldots & x_{k'n} \end{bmatrix}$$

of observations of the independent variables is denoted by the matrix $\mathbf{X}$. We suppose that all these observations are measured from the sample means. The corresponding sets of first differences of the observations are denoted by $\Delta\mathbf{y}$ and $\Delta\mathbf{X}$. The numerator of (2) is the quadratic form

$$(\Delta\mathbf{z})'(\Delta\mathbf{z}) = \mathbf{z}'\mathbf{A}\mathbf{z},$$

where **A** is the real symmetric matrix

$$\begin{bmatrix} 1 & -1 & 0 & \dots & \dots & \dots & 0 \\ -1 & 2 & -1 & \dots & \dots & \dots & \dots \\ 0 & -1 & 2 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & 2 & -1 \\ 0 & \dots & \dots & \dots & 0 & -1 & 1 \end{bmatrix}.$$

The moments of d are obtained by calculating the traces of certain matrices. The trace of a square matrix is simply the sum of the elements in the leading diagonal. For example, the trace of **A**, denoted by tr **A**, is $2(n-1)$, where n is the number of rows or columns in **A**.

It was shown in Part I that the mean and variance of d are given by

$$E(d) = \frac{P}{n-k'-1}, \tag{5}$$

$$\text{var}\,(d) = \frac{2}{(n-k'-1)(n-k'+1)}\{Q - PE(d)\}, \tag{6}$$

where
$$P = \text{tr}\,\mathbf{A} - \text{tr}\,\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}, \tag{7}$$

and
$$Q = \text{tr}\,\mathbf{A}^2 - 2\,\text{tr}\,\{\mathbf{X}'\mathbf{A}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\} + \text{tr}\,[\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}^2]. \tag{8}$$

The elements of $(\mathbf{X}'\mathbf{X})^{-1}$ will have been obtained for the calculation of the regression coefficients, and the elements of $\mathbf{X}'\mathbf{A}\mathbf{X}$ for the calculation of d; for $\mathbf{X}'\mathbf{A}\mathbf{X} = (\mathbf{\Delta X})'(\mathbf{\Delta X})$, so that the (i, j)th element of $\mathbf{X}'\mathbf{A}\mathbf{X}$ is simply the sum of products $\Sigma \Delta x_i \Delta x_j$. Thus the only new matrix requiring calculation is $\mathbf{X}'\mathbf{A}^2\mathbf{X}$. Now $\mathbf{X}'\mathbf{A}^2\mathbf{X}$ is very nearly equal to $(\mathbf{\Delta^2 X})'(\mathbf{\Delta^2 X})$, where $\mathbf{\Delta^2 X}$ represents the matrix of second differences of the independent variables. Thus the (i, j)th element of $\mathbf{X}'\mathbf{A}^2\mathbf{X}$ will usually be given sufficiently closely by $\Sigma(\Delta^2 x_i)(\Delta^2 x_j)$, where $\Delta^2 x_i$ stands for the second difference of the ith independent variable. (More exactly

$$(\mathbf{X}'\mathbf{A}^2\mathbf{X})_{ij} = \Sigma(\Delta^2 x_i)(\Delta^2 x_j) + (x_{i1} - x_{i2})(x_{j1} - x_{j2}) + (x_{in-1} - x_{in})(x_{jn-1} - x_{jn}).)$$

The calculations will be exemplified by the data of Example 1. Referring to § 2 we see that

$$\mathbf{X}'\mathbf{A}\mathbf{X} = \begin{bmatrix} 0{\cdot}023539 & 0{\cdot}000527 \\ 0{\cdot}000527 & 0{\cdot}083559 \end{bmatrix},$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 2{\cdot}966354 & -1{\cdot}014984 \\ -1{\cdot}014984 & 0{\cdot}632006 \end{bmatrix}.$$

Although $\text{tr}\,\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ is simply the sum of the two diagonal elements of the product of these two matrices, we shall need the remaining two elements below. so the whole matrix is computed giving

$$\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0{\cdot}069290 & -0{\cdot}023559 \\ -0{\cdot}083248 & 0{\cdot}052275 \end{bmatrix}.$$

Thus
$$\text{tr}\ \mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = 0{\cdot}069290 + 0{\cdot}052275$$
$$= 0{\cdot}121565.$$

Substituting in (7) and (5) and remembering that $\operatorname{tr}\mathbf{A} = 2(n-1) = 136$ in the present case, we have

$$E(d) = \frac{135{\cdot}878435}{66} = 2{\cdot}05876.$$

The matrix of sums of squares and products of second differences is found to be

$$\mathbf{X}'\mathbf{A}^2\mathbf{X} = \begin{bmatrix} 0{\cdot}035867 & -0{\cdot}004495 \\ -0{\cdot}004495 & 0{\cdot}116368 \end{bmatrix}.$$

$\operatorname{tr}\mathbf{X}'\mathbf{A}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ is obtained by multiplying the first column of $(\mathbf{X}'\mathbf{X})^{-1}$ into the first row of $\mathbf{X}'\mathbf{A}^2\mathbf{X}$ and adding the product of the second column of $(\mathbf{X}'\mathbf{X})^{-1}$ into the second row of $\mathbf{X}'\mathbf{A}^2\mathbf{X}$ giving $\operatorname{tr}\mathbf{X}'\mathbf{A}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = 0{\cdot}189064$. $\operatorname{tr}[\{\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\}^2]$ is simply the sum of squares of the elements of the matrix $\mathbf{X}'\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ i.e. 0·0150189. Also

$$\operatorname{tr}\mathbf{A}^2 = 2(3n-4) = 406.$$

Substituting in (8) and (6) we have

$$\begin{aligned} \operatorname{var} d &= \frac{2}{66\times 68}(405{\cdot}636891 - 135{\cdot}878435 \times 2{\cdot}05876) \\ &= 0{\cdot}0561033. \end{aligned}$$

We now assume that $\frac{1}{4}d$ is distributed in the Beta distribution with density

$$\frac{1}{B(p,q)}\left(\frac{d}{4}\right)^{p-1}\left(1-\frac{d}{4}\right)^{q-1}.$$

This distribution gives

$$E(d) = \frac{4p}{p+q},$$

$$\operatorname{var} d = \frac{16pq}{(p+q)^2(p+q+1)},$$

from which we find p and q by the equations

$$\left.\begin{aligned} p+q &= \frac{E(d)\{4-E(d)\}}{\operatorname{var} d} - 1, \\ p &= \tfrac{1}{4}(p+q)E(d). \end{aligned}\right\} \tag{9}$$

To test against positive serial correlation we require the critical value of $\frac{1}{4}d$ at the lower tail of the distribution. If $2p$ and $2q$ are integers, this can be obtained from Catherine Thompson's tables (1941), or indirectly from tables of the variance ratio or Fisher's z, such as those in the Fisher-Yates tables (1948); if $2p$ and $2q$ are not both integers, a first approximation may be found using the nearest integral values. Thus $F = \dfrac{p(4-d)}{qd}$ is distributed as the variance ratio and $z = \frac{1}{2}\log_e F$ is Fisher's z, both with $n_1 = 2q$, $n_2 = 2p$ degrees of freedom.

For moderately large numbers of observations a convenient way of finding the significance point when $2p$ and $2q$ are not integral is to use Carter's (1947) approximation to Fisher's z. This states that the critical value of z is approximately

$$\frac{\xi\sqrt{(h+\lambda)}}{h} - \left(\frac{1}{2q}-\frac{1}{2p}\right)\left(\lambda+\frac{5}{6}-\frac{s}{3}\right),$$

where

$$s = \frac{1}{2p}+\frac{1}{2q}, \quad h = \frac{2}{s}, \quad \lambda = \frac{\xi^2-3}{6}.$$

The values of ξ and λ to be used for 5 and 1 % tests against positive serial correlation are as follows:

	5%	1%
ξ	1·6449	2·3263
λ	0·0491	0·4020

Returning to the numerical example we find from (9)

$$p = 36{\cdot}1495, \quad q = 34{\cdot}0860, \qquad \text{whence} \quad F = 15{\cdot}99 \quad \text{and} \quad z = 1{\cdot}3848.$$

Carter's approximation gives a critical 1 % value of z of 0·278, which is less than the observed value, thus indicating significant serial correlation. Here the significance is so marked that it may be seen immediately by referring to a table of significant points of z or F around $n_1 = n_2 = 70$ (e.g. Snedecor, 1937).

For testing against negative serial correlation the same procedure is used except that d is replaced throughout by $4-d$.

5. One- and two-way classifications

In any regression analysis where the independent variables assume the same values in all applications it would be theoretically possible to dispense with the bounds and tabulate the significance points of d once and for all. This could be done, for instance, for analysis of variance models such as one- and two-way classifications and for polynomial regressions with equally spaced variate values. The calculation of tables of this type is rather a formidable task and only one set has so far been published, namely, the significance points of the circular serial correlation coefficient of the least squares residuals from Fourier regressions, tabulated by R. L. & T. W. Anderson (1950). Pending the publication of further tables the bounds test of the present paper may be used, with the approximate procedure described in § 4 for cases in which the bounds do not give a decisive result, or when the n and k' are beyond the range of Tables 4, 5 and 6. In the present section the calculation of d is described for one- and two-way classifications and expressions are given for its mean and variance, while in the next section the same is done for polynomial regressions.

It is convenient to think of the observations as a time series consisting of monthly observations recorded for a number of years, though the results are of course of general application. We may fit constants for years only or for months only or for both months and years. An exact test of serial correlation in the 'months only' case may be made by means of R. L. & T. W. Anderson's tables (1950). For the 'years only' and the 'months and years' case no exact test is at present available. The tests described in this paper may, however, be used in all three cases.

If there are s 'years' each of t 'months', the mean and variance of d for the three models are as follows:

'Years' only:

$$\left.\begin{aligned} E(d) &= 2\left(1+\frac{1}{t}-\frac{1}{st}\right), \\ \operatorname{var} d &= \frac{4}{s(t-1)(st-s+2)}\left\{st-2s-\frac{2s}{t}+\frac{4}{t}+\frac{5s}{t^2}-\frac{8}{t^2}-\frac{2}{st}+\frac{2}{st^2}\right\}. \end{aligned}\right\} \tag{10}$$

'Months' only:

$$\left.\begin{aligned} E(d) &= 2\left(1-\frac{1}{st}\right), \\ \operatorname{var} d &= \frac{4}{t(s-1)(st-t+2)}\left\{st-t-1+\frac{1}{s^2}-\frac{2}{st}+\frac{2}{s^2t}\right\}. \end{aligned}\right\} \tag{11}$$

'Years and months':

$$\left.\begin{aligned} E(d) &= 2\left\{1+\frac{1}{t}-\frac{1}{s(t-1)}\right\}, \\ \operatorname{var} d &= \frac{4}{(s-1)(t-1)^2(st-t-s+3)} \\ &\quad \left\{st^2-t^2-3st+3t+4-\frac{2t}{s}+\frac{7s}{t}-\frac{12}{t}-\frac{5s}{t^2}+\frac{2t}{s^2}+\frac{6}{t^2}\right\}. \end{aligned}\right\} \tag{12}$$

These formulae were found by substituting the appropriate matrices into the general formulae given in Part I. The subsequent reductions were straightforward but extremely tedious.

Example 2. To illustrate the test procedures for models of this kind we shall consider the two-way classification of the data in Table 2 on the receipts of butter (in units of 1,000,000 lb.) at five markets (Boston, Chicago, San Francisco, Milwaukee and St Louis). The same data, for 1935, 1936 and 1937 only, were used by R. L. & T. W. Anderson (1950) for illustrating their procedure for testing serial correlation in the 'months only' case. The figures in parentheses are residuals from the monthly averages.

Table 2. *Receipts of butter (millions of lb. weight) at five U.S.A. markets*

Month	Year					Average
	1933	1934	1935	1936	1937	
Jan.	58·3 (+8·20)	52·6 (+2·50)	48·9 (−1·20)	48·3 (−1·80)	42·4 (−7·70)	50·10
Feb.	51·3 (+5·28)	46·9 (+0·88)	43·4 (−2·62)	47·1 (+1·08)	41·4 (−4·62)	46·02
March	58·1 (+5·86)	57·9 (+5·66)	43·8 (−8·44)	52·4 (+0·16)	49·0 (−3·24)	52·24
April	55·1 (+1·86)	54·2 (+0·96)	50·8 (−2·44)	55·3 (+2·06)	50·8 (−2·44)	53·24
May	74·6 (+5·94)	70·6 (+1·94)	67·6 (−1·06)	64·7 (−3·96)	65·8 (−2·86)	68·66
June	83·9 (+2·64)	73·3 (−7·96)	83·7 (+2·44)	79·5 (−1·76)	85·9 (+4·64)	81·26
July	73·5 (+1·56)	70·3 (−1·64)	82·7 (+10·76)	62·6 (−9·34)	70·6 (−1·34)	71·94
Aug.	73·3 (+11·78)	66·4 (+4·88)	60·8 (−0·72)	51·3 (−10·22)	55·8 (−5·72)	61·52
Sept.	63·0 (+7·96)	56·7 (+1·66)	55·4 (+0·36)	51·0 (−4·04)	49·1 (−5·94)	55·04
Oct.	58·3 (+5·58)	57·2 (+4·48)	48·4 (−4·32)	54·0 (+1·28)	45·7 (−7·02)	52·72
Nov.	55·1 (+9·20)	47·7 (+1·80)	37·7 (−8·20)	45·2 (−0·70)	43·8 (−2·10)	45·90
Dec.	56·5 (+9·70)	44·9 (−1·90)	41·0 (−5·80)	44·9 (−1·90)	46·7 (−0·10)	46·80
Average	63·42	58·22	55·35	54·69	53·92	57·12
Total	761·0	698·7	664·2	656·3	647·0	3427·2

Source: *Agricultural Statistics*, United States Government Printing Office, Washington, D.C., 1939, p. 390.

We require to test for serial correlation in the model

$$y_{ij} = \mu+\alpha_i+\beta_j+\epsilon_{ij} \quad (i=1,2,\ldots,5;\ j=1,2,\ldots,12),$$

where y_{ij} is the observation in the jth month of the ith year, μ, α_i and β_j are constants and ϵ_{ij} is the error term.

The least squares estimates of μ, α_i and β_j are m, $a_i - m$ and $b_j - m$, where m is the sample mean of all the observations and a_i and b_j are the means of observations in the ith year and the jth month respectively. Thus the residuals are given by

$$z_{ij} = y_{ij} - a_i - b_j + m.$$

The test is made as before by calculating

$$d = \frac{\Sigma(\Delta z_{ij})^2}{\Sigma z_{ij}^2},$$

where the Δz_{ij}'s are the first differences of the residuals when arranged as a single time series.

$\Sigma(\Delta z)^2$ may be calculated by working out the individual residuals from the monthly averages and finding their first differences. The difference between the December value of one year and the January value of the succeeding year then needs to be adjusted to take account of the difference between the two yearly averages. For instance, the difference between the 'months only' residuals for December 1933 and January 1934 is

$$2{\cdot}50 - 9{\cdot}70 = -7{\cdot}20.$$

From this must be subtracted the difference between the 1934 and 1933 yearly averages, i.e. $58{\cdot}22 - 63{\cdot}42 = -5{\cdot}20$. The net difference is therefore $-7{\cdot}20 + 5{\cdot}20 = -2{\cdot}00$. The sum of the resulting differences squared is $\Sigma(\Delta z)^2$. For the calculation of Σz^2 the normal method for the residual sum of squares may be used, i.e. find the sum of squares of the original observations and subtract the sums of squares due to the fitted constants. Alternatively, with the above method of calculating $\Sigma(\Delta z)^2$, the sum of squares of the residuals from the monthly averages may be calculated directly, from which it only remains to subtract the sum of squares due to years.

For the data in the table we find $\Sigma(\Delta z)^2 = 1191{\cdot}2454$ and $\Sigma z^2 = 850{\cdot}8890$, giving $d = \dfrac{1191{\cdot}2454}{850{\cdot}8890} = 1{\cdot}4000$. It remains to test the significance of this value of d by the method of § 4. For this purpose the formulae (12) may be evaluated with $s = 5$ and $t = 12$ to give

$$E(d) = 2{\cdot}1303, \quad \operatorname{var} d = 0{\cdot}077964.$$

If $\frac{1}{4}d$ is assumed to have a Beta-distribution with parameters p, q, then values of $E(d)$ and $\operatorname{var} d$ may be substituted in the formulae (9) to give

$$p = 26{\cdot}6758, \quad q = 23{\cdot}4125.$$

The observed value of d is 1·4000, so that

$$F = \frac{p(4-d)}{pd} = 2{\cdot}16,$$

with $n_1 \sim 47$, $n_2 \sim 53$ degrees of freedom. A cursory examination of the significant points of F shows that the 5 % point is certainly less than 1·63, while the 1 % point is less than 2·00. Thus our value of F is significant at the 1 % level, and the hypothesis of serial independence of the errors in the above model may be considered untenable.

6. Polynomial regressions

An important application of the least squares method in time-series analysis is in the fitting of polynomial trend lines. When the values of the time variate (or its equivalent in other applications) are spaced at equal intervals, the fitting is carried out most expeditiously by

means of orthogonal polynomials. In what follows we shall assume that the ξ' polynomials tabulated by Fisher & Yates (1948), and in a more extended form by Anderson & Houseman (1942), have been used. The regression model is

$$y = \beta_0 + \beta_1 \xi_1' + \beta_2 \xi_2' + \ldots + \beta_k \xi_k' + \epsilon,$$

where ξ_i' is the polynomial of ith degree in x, the independent variable, which we suppose takes the values $1, 2, \ldots, n$, $\beta_0, \beta_1, \ldots, \beta_k$, are constants and ϵ is the error term. We require to test the error term for serial correlation.

The test procedure is a good deal less laborious in this application than in the ordinary regression case described in § 2. We shall illustrate it by considering the following example. The data were taken by Anderson & Houseman (1942) from Schultz's demand studies (1938) to illustrate the routine procedure of fitting the polynomials. We shall not, therefore, give details of the initial calculations but refer the reader instead to Anderson & Houseman's bulletin.

Table 3. *Price of sugar*, 1875–1936

Year	Price	Year	Price	Year	Price	Year	Price
1875	67	1891	6	1907	6	1923	44
1876	65	1892	3	1908	10	1924	35
1877	73	1893	8	1909	8	1925	15
1878	55	1894	1	1910	10	1926	15
1879	48	1895	2	1911	13	1927	18
1880	56	1896	5	1912	10	1928	15
1881	57	1897	5	1913	3	1929	10
1882	52	1898	10	1914	7	1930	6
1883	45	1899	9	1915	16	1931	4
1884	28	1900	13	1916	29	1932	0
1885	24	1901	10	1917	37	1933	3
1886	21	1902	5	1918	38	1934	1
1887	20	1903	6	1919	50	1935	3
1888	30	1904	8	1920	74	1936	7
1889	36	1905	13	1921	22		
1890	22	1906	5	1922	19		

Example 3. A polynomial time trend is to be fitted to 62 annual sugar prices (1875–1936) given in Table 3. The prices are in terms of mills (tenths of a cent) coded by subtracting 40 mills from each price.

Anderson & Houseman find the following values for the sums of squares and products, giving the regression coefficients shown:

$$\Sigma y = 1{,}336 \qquad b_0 = 21{\cdot}548$$
$$\Sigma(y-\bar{y})^2 = 25{,}250$$
$$\Sigma(\xi_1')^2 = 79{,}422 \qquad \Sigma(\xi_2')^2 = 1{,}270{,}752$$
$$\Sigma(\xi_3')^2 = 139{,}238{,}112 \qquad \Sigma(\xi_4')^2 = 103{,}639{,}568{,}032$$
$$\Sigma(y\xi_1') = -20{,}286 \qquad b_1 = -0{\cdot}2554$$
$$\Sigma(y\xi_2') = 72{,}775 \qquad b_2 = 0{\cdot}05727$$
$$\Sigma(y\xi_3') = -1{,}080{,}557 \qquad b_3 = -0{\cdot}0077605$$
$$\Sigma(y\xi_4') = -7{,}599{,}201 \qquad b_4 = -0{\cdot}00007332$$

To test for serial correlation we must calculate

$$d = \frac{\Sigma(\Delta z)^2}{\Sigma z^2}.$$

Taking first the case in which terms only as far as ξ_3' are fitted,

$$z = y - b_0 - b_1\xi_1' - b_2\xi_2' - b_3\xi_3'$$

and

$$\Sigma z^2 = 25{,}250 - (-0{\cdot}2554)(-20{,}286) - (0{\cdot}05727)(72{,}775) - (-0{\cdot}0077605)(-1{,}080{,}557)$$
$$= 7515, \tag{13}$$

$$\Sigma(\Delta z)^2 = \Sigma(\Delta y)^2 - 2b_1\Sigma\Delta y\Delta\xi_1' - 2b_2\Sigma\Delta y\Delta\xi_2' - 2b_3\Sigma\Delta y\Delta\xi_3' + \sum_{i=1}^{3}\sum_{j=1}^{3} b_i b_j \Sigma\Delta\xi_i'\Delta\xi_j'. \tag{14}$$

$\Sigma(\Delta y)^2$ is calculated directly as the sum of the squares of the first differences of the series of observations of y; its value here is 2590. For the remaining terms, indirect methods are much quicker. It may be verified that

$$\left.\begin{aligned}
\Sigma\Delta y\Delta\xi_1' &= (y_n - y_1)\,\Delta\xi_1'(0),\\
\Sigma\Delta y\Delta\xi_2' &= -2\lambda_2\Sigma y - (y_n + y_1)\,\Delta\xi_2'(0),\\
\Sigma\Delta y\Delta\xi_3' &= -\frac{6\lambda_3}{\lambda_1}\Sigma y\xi_1' + (y_n - y_1)\,\Delta\xi_3'(0),\\
\Sigma\Delta y\Delta\xi_4' &= -\lambda_4\left[\frac{12}{\lambda_2}\Sigma y\xi_2' + \left\{\frac{1}{\lambda_2}(n^2-1) - \frac{3n^2-13}{7}\right\}\Sigma y\right] - (y_n + y_1)\,\Delta\xi_4'(0),\\
\Sigma\Delta y\Delta\xi_5' &= \lambda_5\left[\frac{20}{\lambda_3}\Sigma y\xi_3' + \left\{\frac{1}{\lambda_3}(3n^2-7) - \frac{10}{\lambda_1}\frac{n^2-13}{3}\right\}\Sigma y\xi_1'\right] + (y_n - y_1)\,\Delta\xi_5'(0),
\end{aligned}\right\} \tag{15}$$

and that, with $(i \leqslant j)$,

$$\left.\begin{aligned}
\Sigma\Delta\xi_i'\Delta\xi_j' &= -2\xi_j'(1)\,\Delta\xi_i'(0) \quad (i+j \text{ even}),\\
&= 0 \qquad\qquad\quad (i+j \text{ odd}).
\end{aligned}\right\} \tag{16}$$

In these expressions it is assumed that the original time variate x takes the values $1, 2, \ldots, n$. $\xi_j'(1)$ denotes the value of ξ_j' for $x = 1$. Similarly, $\Delta\xi_j'(0)$ denotes the value of $\Delta\xi_j'$ for $x = 0$, i.e. $\xi_j'(1) - \xi_j'(0)$. y_1 and y_n are the first and last observations in the series of values of the dependent variate. The λ_i is given for each n at the foot of the appropriate column of ξ_i' values in the published tables.

The values of $\Delta\xi_i'(0)$ are obtained by writing down the first few terms of the series $\xi_i'(x)$ for $x = 1, 2, \ldots$, and preparing a small table of differences. $\Delta\xi_i'(0)$ is then found by simple addition. It should be noted that the values of $\xi_i'(x)$ should be read upwards starting at the bottom of the published table, and that the signs should be reversed for polynomials of odd degree. Example 3, with $n = 62$, gives the following lay-out. The values of $\Delta\xi_i'(0)$ required are printed in italics:

x	ξ_1'	$\Delta\xi_1'$	x	ξ_2'	$\Delta\xi_2'$	$\Delta^2\xi_2'$	x	ξ_3'	$\Delta\xi_3'$	$\Delta^2\xi_3'$	$\Delta^3\xi_3'$
		2			*−31*				*769*		
1	−61	2	1	305	−30	1	1	−3599	708	−61	2
2	−59	2	2	275	−29	1	2	−2891	649	−59	2
3	−57		3	246	−28		3	−2242	592	−57	2
			4	218			4	−1650	437	−55	
							5	−1113			

Substituting in formulae (15, 16) we have

$$\begin{aligned}
\Sigma\Delta y\Delta\xi_1' &= (7-67)(2) = -120,\\
\Sigma\Delta y\Delta\xi_2' &= (-2)(\tfrac{1}{2})(1336)-(7+67)(-31) = 958,\\
\Sigma\Delta y\Delta\xi_3' &= -\frac{6}{2\cdot 3}(-20{,}286)+(7-67)769 = -25{,}854,\\
\Sigma(\Delta\xi_1')^2 &= -2(-61)(2) = 244,\\
\Sigma(\Delta\xi_2')^2 &= -2(305)(-31) = 18{,}910,\\
\Sigma(\Delta\xi_3')^2 &= -2(-3599)(769) = 5{,}535{,}262,\\
\Sigma\Delta\xi_1'\Delta\xi_2' &= \Sigma\Delta\xi_2'\Delta\xi_3' = 0,\\
\Sigma\Delta\xi_1'\Delta\xi_3' &= -2(-3599)(2) = 14{,}396.
\end{aligned}$$

We now have all the quantities necessary for the calculation of $\Sigma(\Delta z)^2$. Substituting in (14) we have

$$\begin{aligned}
\Sigma(\Delta z)^2 &= 2590-2(-0{\cdot}2554)(120)-2(0{\cdot}05727)(958)\\
&\quad -2(-0{\cdot}0077605)(-25{,}854)+2(-0{\cdot}2554)(-0{\cdot}0077605)(14{,}396)\\
&\quad +(0{\cdot}2554)^2(244)+(0{\cdot}05727)^2(18{,}910)+(0{\cdot}0077605)^2(5{,}535{,}262)\\
&= 2486{\cdot}06.
\end{aligned}$$

Thus
$$d = \frac{2486{\cdot}06}{7515} = 0{\cdot}3308.$$

Reference to Table 6 shows that this value is significant at the 1 % level and therefore provides significant evidence of the existence of positive serial correlation.

When a polynomial of the fourth degree is fitted to the data a value of d of 0·3603 is obtained. This still remains highly significant at the 1 % level.

Mean and variance of d

The quickest way of calculating the mean and variance of d for polynomial regressions is to use the numerical procedure described below. It is, however, possible to obtain explicit formulae, and these have been calculated for polynomials up to the fifth degree. Owing to the complexity of the resulting expressions we shall only give them for $k' = 1, 2, 3$.

Taking first the numerical procedure, let $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ denote the column vectors of values of the polynomials (the usual prime is omitted in the vector form since we wish to use the same sign for the matrix operation of transposition). The mean and variance of d are given by (5) and (6) with

$$P = 2(n-1)-\sum_{i=1}^{k'}\frac{\boldsymbol{\xi}_i'\mathbf{A}\boldsymbol{\xi}_i}{\boldsymbol{\xi}_i'\boldsymbol{\xi}_i},$$

$$Q = 2(3n-4)-2\sum_{i=1}^{k'}\frac{\boldsymbol{\xi}_i'\mathbf{A}^2\boldsymbol{\xi}_i}{\boldsymbol{\xi}_i'\boldsymbol{\xi}_i}+\sum_{i=1}^{k'}\left(\frac{\boldsymbol{\xi}_i'\mathbf{A}\boldsymbol{\xi}_i}{\boldsymbol{\xi}_i'\boldsymbol{\xi}_i}\right)^2+2\sum_{j=2}^{k'}\sum_{i=1}^{j-1}\frac{(\boldsymbol{\xi}_i'\mathbf{A}\boldsymbol{\xi}_j)^2}{\boldsymbol{\xi}_i'\boldsymbol{\xi}_i\boldsymbol{\xi}_j'\boldsymbol{\xi}_j},$$

as was shown in Part I, $\mathbf{A}$ being the matrix defined in § 4. The quantities $\boldsymbol{\xi}_i'\boldsymbol{\xi}_i$ are the sums of squares of the values of ξ_i', given at the foot of the ξ' tables. The quantities $\boldsymbol{\xi}_i'\mathbf{A}\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}_i'\mathbf{A}^2\boldsymbol{\xi}_i$ may be found from the expressions

$$\boldsymbol{\xi}_i'\mathbf{A}\boldsymbol{\xi}_j = \begin{cases} -2\xi_j'(1)\,\Delta\xi_i'(0) & (i+j \text{ even}),\\ 0 & (i+j \text{ odd}),\end{cases}$$

$$\boldsymbol{\xi}_i'\mathbf{A}^2\boldsymbol{\xi}_i = 2\xi_i'(1)\,\Delta^3\xi_i'(-1)+2\Delta\xi_i'(0)\,\Delta\xi_i'(1).$$

The values of $\xi_i'(x)$ and its differences may easily be found from the published tables.

Applying the method to Example 3 we find

$$\begin{aligned} P &= 122-\frac{244}{79{,}422}-\frac{18{,}910}{1{,}270{,}752}-\frac{5{,}535{,}262}{139{,}238{,}112} \\ &= 121{\cdot}9423, \\ Q &= 368-2\left(\frac{8}{79{,}422}+\frac{1860}{1{,}270{,}752}+\frac{1{,}074{,}508}{139{,}238{,}112}\right) \\ &\quad +\left(\frac{244}{79{,}422}\right)^2+\left(\frac{18{,}910}{1{,}270{,}752}\right)^2+\left(\frac{5{,}535{,}262}{139{,}238{,}112}\right)^2 \\ &\quad +2\left(\frac{14{,}396}{79{,}422\times 139{,}238{,}112}\right) \\ &= 367{\cdot}9832. \end{aligned}$$

The explicit formulae for P and Q are as follows:

$$P = 2\left(n-1-\sum_{i=1}^{k'} p_i\right),$$

where $$p_1=\frac{6}{n(n+1)}, \quad p_2=\frac{30}{(n+1)(n+2)}, \quad p_3=\frac{84(n^2+1)}{n(n+1)(n+2)(n+3)},$$

and $$Q = 2\left(3n-4+\sum_{i=1}^{k'} q_i\right),$$

where $$q_1=2p_1^2-\frac{24}{n(n^2-1)}, \quad q_2=2p_2^2-\frac{360}{(n^2-1)(n+2)},$$

$$q_3=2p_3^2+\frac{336(n-2)(n-3)}{n^2(n+1)^2(n+2)(n+3)}-\frac{336(3n^2+10n-7)}{n(n^2-1)(n+2)(n+3)}.$$

Appendix on the calculation of the tables

1. *Tables* 4, 5 *and* 6. The exact distribitions of d_L and d_U, whose significance points are required for Tables 4, 5 and 6, are not known. When transformed to the range (0, 1), their probability densities may, however, be represented by the series

$$g(x)=\frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}\left\{1+\sum_{s=1}^{\infty} a_s G_s(x)\right\}, \tag{17}$$

where the G's are the polynomials (Jacobi) which are orthogonal on the range (0, 1) with respect to the weight function $\frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}$. These polynomials are defined by (see Courant & Hilbert,* 1931)

$$\begin{aligned} G_s(x) &= 1-\frac{p+q-1+s}{p}x+\frac{(p+q-1+s)(p+q+s)}{p(p+1)}x^2-\dots \\ &\quad +(-1)^s\frac{(p+q-1+s)(p+q+s)\dots(p+q+2-2s)}{p(p+1)\dots(p+s-1)}x^s. \end{aligned}$$

The coefficients a_s may be determined by the method of moments; the distribution of d_L and d_U is then a series of Incomplete Beta Functions.

* Note, however, the misprint: $x^q(1-x)^{p-q}$ should read $x^{q-1}(1-x)^{p-q}$.

The weight function was chosen to be the density of the Beta-distribution with the correct mean and variance. (An alternative weight function giving the right order of vanishing of (17) at $x = 0$ and $x = 1$ was also tried but it was found to be less satisfactory.) With this weight function, the coefficients a_1 and a_2 in (17) were zero. Terms as far as $G_4(x)$ were used.

Table 4. *Significance points of d_L and d_U: 5 %*

n	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1·08	1·36	0·95	1·54	0·82	1·75	0·69	1·97	0·56	2·21
16	1·10	1·37	0·98	1·54	0·86	1·73	0·74	1·93	0·62	2·15
17	1·13	1·38	1·02	1·54	0·90	1·71	0·78	1·90	0·67	2·10
18	1·16	1·39	1·05	1·53	0·93	1·69	0·82	1·87	0·71	2·06
19	1·18	1·40	1·08	1·53	0·97	1·68	0·86	1·85	0·75	2·02
20	1·20	1·41	1·10	1·54	1·00	1·68	0·90	1·83	0·79	1·99
21	1·22	1·42	1·13	1·54	1·03	1·67	0·93	1·81	0·83	1·96
22	1·24	1·43	1·15	1·54	1·05	1·66	0·96	1·80	0·86	1·94
23	1·26	1·44	1·17	1·54	1·08	1·66	0·99	1·79	0·90	1·92
24	1·27	1·45	1·19	1·55	1·10	1·66	1·01	1·78	0·93	1·90
25	1·29	1·45	1·21	1·55	1·12	1·66	1·04	1·77	0·95	1·89
26	1·30	1·46	1·22	1·55	1·14	1·65	1·06	1·76	0·98	1·88
27	1·32	1·47	1·24	1·56	1·16	1·65	1·08	1·76	1·01	1·86
28	1·33	1·48	1·26	1·56	1·18	1·65	1·10	1·75	1·03	1·85
29	1·34	1·48	1·27	1·56	1·20	1·65	1·12	1·74	1·05	1·84
30	1·35	1·49	1·28	1·57	1·21	1·65	1·14	1·74	1·07	1·83
31	1·36	1·50	1·30	1·57	1·23	1·65	1·16	1·74	1·09	1·83
32	1·37	1·50	1·31	1·57	1·24	1·65	1·18	1·73	1·11	1·82
33	1·38	1·51	1·32	1·58	1·26	1·65	1·19	1·73	1·13	1·81
34	1·39	1·51	1·33	1·58	1·27	1·65	1·21	1·73	1·15	1·81
35	1·40	1·52	1·34	1·58	1·28	1·65	1·22	1·73	1·16	1·80
36	1·41	1·52	1·35	1·59	1·29	1·65	1·24	1·73	1·18	1·80
37	1·42	1·53	1·36	1·59	1·31	1·66	1·25	1·72	1·19	1·80
38	1·43	1·54	1·37	1·59	1·32	1·66	1·26	1·72	1·21	1·79
39	1·43	1·54	1·38	1·60	1·33	1·66	1·27	1·72	1·22	1·79
40	1·44	1·54	1·39	1·60	1·34	1·66	1·29	1·72	1·23	1·79
45	1·48	1·57	1·43	1·62	1·38	1·67	1·34	1·72	1·29	1·78
50	1·50	1·59	1·46	1·63	1·42	1·67	1·38	1·72	1·34	1·77
55	1·53	1·60	1·49	1·64	1·45	1·68	1·41	1·72	1·38	1·77
60	1·55	1·62	1·51	1·65	1·48	1·69	1·44	1·73	1·41	1·77
65	1·57	1·63	1·54	1·66	1·50	1·70	1·47	1·73	1·44	1·77
70	1·58	1·64	1·55	1·67	1·52	1·70	1·49	1·74	1·46	1·77
75	1·60	1·65	1·57	1·68	1·54	1·71	1·51	1·74	1·49	1·77
80	1·61	1·66	1·59	1·69	1·56	1·72	1·53	1·74	1·51	1·77
85	1·62	1·67	1·60	1·70	1·57	1·72	1·55	1·75	1·52	1·77
90	1·63	1·68	1·61	1·70	1·59	1·73	1·57	1·75	1·54	1·78
95	1·64	1·69	1·62	1·71	1·60	1·73	1·58	1·75	1·56	1·78
100	1·65	1·69	1·63	1·72	1·61	1·74	1·59	1·76	1·57	1·78

A first set of significance points was obtained using the weight function as a first approximation; these values were then adjusted using the higher terms of the series. The first set of values was calculated partly by Wise's* (1950) method and partly by means of Carter's (1947) approximation. The adjustments necessary were found to be small and to vary very slowly with p and q; they could therefore be calculated by the following method which reduces to a minimum interpolation in the *Tables of the Incomplete Beta Function* (1948).

* We are indebted to Mr Wise for some helpful correspondence on his method.

First, p and q may be replaced by the integers nearest them. For these integers an exact significance point may be found by quadratic inverse interpolation. The difference of this exact point and the first approximation to it is the required adjustment. The adjustment required for the fourth moment turned out to be negligible to the order of accuracy aimed at and could have been ignored. The adjustments were so small and regular that it was only

Table 5. *Significance points of d_L and d_U*: 2·5 %

n	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0·95	1·23	0·83	1·40	0·71	1·61	0·59	1·84	0·48	2·09
16	0·98	1·24	0·86	1·40	0·75	1·59	0·64	1·80	0·53	2·03
17	1·01	1·25	0·90	1·40	0·79	1·58	0·68	1·77	0·57	1·98
18	1·03	1·26	0·93	1·40	0·82	1·56	0·72	1·74	0·62	1·93
19	1·06	1·28	0·96	1·41	0·86	1·55	0·76	1·72	0·66	1·90
20	1·08	1·28	0·99	1·41	0·89	1·55	0·79	1·70	0·70	1·87
21	1·10	1·30	1·01	1·41	0·92	1·54	0·83	1·69	0·73	1·84
22	1·12	1·31	1·04	1·42	0·95	1·54	0·86	1·68	0·77	1·82
23	1·14	1·32	1·06	1·42	0·97	1·54	0·89	1·67	0·80	1·80
24	1·16	1·33	1·08	1·43	1·00	1·54	0·91	1·66	0·83	1·79
25	1·18	1·34	1·10	1·43	1·02	1·54	0·94	1·65	0·86	1·77
26	1·19	1·35	1·12	1·44	1·04	1·54	0·96	1·65	0·88	1·76
27	1·21	1·36	1·13	1·44	1·06	1·54	0·99	1·64	0·91	1·75
28	1·22	1·37	1·15	1·45	1·08	1·54	1·01	1·64	0·93	1·74
29	1·24	1·38	1·17	1·45	1·10	1·54	1·03	1·63	0·96	1·73
30	1·25	1·38	1·18	1·46	1·12	1·54	1·05	1·63	0·98	1·73
31	1·26	1·39	1·20	1·47	1·13	1·55	1·07	1·63	1·00	1·72
32	1·27	1·40	1·21	1·47	1·15	1·55	1·08	1·63	1·02	1·71
33	1·28	1·41	1·22	1·48	1·16	1·55	1·10	1·63	1·04	1·71
34	1·29	1·41	1·24	1·48	1·17	1·55	1·12	1·63	1·06	1·70
35	1·30	1·42	1·25	1·48	1·19	1·55	1·13	1·63	1·07	1·70
36	1·31	1·43	1·26	1·49	1·20	1·56	1·15	1·63	1·09	1·70
37	1·32	1·43	1·27	1·49	1·21	1·56	1·16	1·62	1·10	1·70
38	1·33	1·44	1·28	1·50	1·23	1·56	1·17	1·62	1·12	1·70
39	1·34	1·44	1·29	1·50	1·24	1·56	1·19	1·63	1·13	1·69
40	1·35	1·45	1·30	1·51	1·25	1·57	1·20	1·63	1·15	1·69
45	1·39	1·48	1·34	1·53	1·30	1·58	1·25	1·63	1·21	1·69
50	1·42	1·50	1·38	1·54	1·34	1·59	1·30	1·64	1·26	1·69
55	1·45	1·52	1·41	1·56	1·37	1·60	1·33	1·64	1·30	1·69
60	1·47	1·54	1·44	1·57	1·40	1·61	1·37	1·65	1·33	1·69
65	1·49	1·55	1·46	1·59	1·43	1·62	1·40	1·66	1·36	1·69
70	1·51	1·57	1·48	1·60	1·45	1·63	1·42	1·66	1·39	1·70
75	1·53	1·58	1·50	1·61	1·47	1·64	1·45	1·67	1·42	1·70
80	1·54	1·59	1·52	1·62	1·49	1·65	1·47	1·67	1·44	1·70
85	1·56	1·60	1·53	1·63	1·51	1·65	1·49	1·68	1·46	1·71
90	1·57	1·61	1·55	1·64	1·53	1·66	1·50	1·69	1·48	1·71
95	1·58	1·62	1·56	1·65	1·54	1·67	1·52	1·69	1·50	1·71
100	1·59	1·63	1·57	1·65	1·55	1·67	1·53	1·70	1·51	1·72

necessary to calculate them at 39 places in the entire set of tables, the remainder being obtained by linear interpolation. The adjustments were negligible for numbers of observations greater than 40.

As a partial check on the calculating procedure it was applied to the calculation of the significance points for a related distribution for which exact significance points were available. To make the circumstances as unfavourable as possible the case $n = 16$, $k' = 5$ at the

Table 6. *Significance points of d_L and d_U: 1 %*

	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0·81	1·07	0·70	1·25	0·59	1·46	0·49	1·70	0·39	1·96
16	0·84	1·09	0·74	1·25	0·63	1·44	0·53	1·66	0·44	1·90
17	0·87	1·10	0·77	1·25	0·67	1·43	0·57	1·63	0·48	1·85
18	0·90	1·12	0·80	1·26	0·71	1·42	0·61	1·60	0·52	1·80
19	0·93	1·13	0·83	1·26	0·74	1·41	0·65	1·58	0·56	1·77
20	0·95	1·15	0·86	1·27	0·77	1·41	0·68	1·57	0·60	1·74
21	0·97	1·16	0·89	1·27	0·80	1·41	0·72	1·55	0·63	1·71
22	1·00	1·17	0·91	1·28	0·83	1·40	0·75	1·54	0·66	1·69
23	1·02	1·19	0·94	1·29	0·86	1·40	0·77	1·53	0·70	1·67
24	1·04	1·20	0·96	1·30	0·88	1·41	0·80	1·53	0·72	1·66
25	1·05	1·21	0·98	1·30	0·90	1·41	0·83	1·52	0·75	1·65
26	1·07	1·22	1·00	1·31	0·93	1·41	0·85	1·52	0·78	1·64
27	1·09	1·23	1·02	1·32	0·95	1·41	0·88	1·51	0·81	1·63
28	1·10	1·24	1·04	1·32	0·97	1·41	0·90	1·51	0·83	1·62
29	1·12	1·25	1·05	1·33	0·99	1·42	0·92	1·51	0·85	1·61
30	1·13	1·26	1·07	1·34	1·01	1·42	0·94	1·51	0·88	1·61
31	1·15	1·27	1·08	1·34	1·02	1·42	0·96	1·51	0·90	1·60
32	1·16	1·28	1·10	1·35	1·04	1·43	0·98	1·51	0·92	1·60
33	1·17	1·29	1·11	1·36	1·05	1·43	1·00	1·51	0·94	1·59
34	1·18	1·30	1·13	1·36	1·07	1·43	1·01	1·51	0·95	1·59
35	1·19	1·31	1·14	1·37	1·08	1·44	1·03	1·51	0·97	1·59
36	1·21	1·32	1·15	1·38	1·10	1·44	1·04	1·51	0·99	1·59
37	1·22	1·32	1·16	1·38	1·11	1·45	1·06	1·51	1·00	1·59
38	1·23	1·33	1·18	1·39	1·12	1·45	1·07	1·52	1·02	1·58
39	1·24	1·34	1·19	1·39	1·14	1·45	1·09	1·52	1·03	1·58
40	1·25	1·34	1·20	1·40	1·15	1·46	1·10	1·52	1·05	1·58
45	1·29	1·38	1·24	1·42	1·20	1·48	1·16	1·53	1·11	1·58
50	1·32	1·40	1·28	1·45	1·24	1·49	1·20	1·54	1·16	1·59
55	1·36	1·43	1·32	1·47	1·28	1·51	1·25	1·55	1·21	1·59
60	1·38	1·45	1·35	1·48	1·32	1·52	1·28	1·56	1·25	1·60
65	1·41	1·47	1·38	1·50	1·35	1·53	1·31	1·57	1·28	1·61
70	1·43	1·49	1·40	1·52	1·37	1·55	1·34	1·58	1·31	1·61
75	1·45	1·50	1·42	1·53	1·39	1·56	1·37	1·59	1·34	1·62
80	1·47	1·52	1·44	1·54	1·42	1·57	1·39	1·60	1·36	1·62
85	1·48	1·53	1·46	1·55	1·43	1·58	1·41	1·60	1·39	1·63
90	1·50	1·54	1·47	1·56	1·45	1·59	1·43	1·61	1·41	1·64
95	1·51	1·55	1·49	1·57	1·47	1·60	1·45	1·62	1·42	1·64
100	1·52	1·56	1·50	1·58	1·48	1·60	1·46	1·63	1·44	1·65

extreme of the tabulated values was examined. The distributions of d_L and d_U for these values were modified to give latent roots occurring in equal pairs, i.e. five pairs in all. The new roots were chosen to lie midway between the roots of the original distribution, thus preserving its asymmetry and general character. By pairing the roots in this way the exact significance points could be determined using results given by R. L. Anderson (1942). The significance points obtained by the approximate procedure agreed with these exact significance points to the order of accuracy required here.

7. AN EXACT BOUNDS TEST

We have given elsewhere (Watson & Durbin, 1951) a general method of constructing exact tests of serial independence which do not require the use of circular definitions of the serial correlation coefficient. The method can be used to obtain the exact distributions of bounding

statistics similar to d_L and d_U. The advantage of having exact distributions is obtained at the cost of throwing away a certain amount of relevant information.

For testing the independence of the errors in a regression model a statistic d' is defined which is a slight modification of d. If the number of observations is even, $2m$ say, then

$$d' = \frac{(z_1 - z_2)^2 + \ldots + (z_{m-1} - z_m)^2 + (z_{m+1} - z_{m+2})^2 + \ldots + (z_{2m-1} - z_{2m})^2}{\sum_1^{2m} z_i^2},$$

and if it is odd, $2m+1$ say, then

$$d' = \frac{(z_1 - z_2)^2 + \ldots + (z_{m-1} - z_m)^2 + (z_{m+2} - z_{m+3})^2 + \ldots + (z_{2m} - z_{2m+1})^2}{\sum_1^{2m+1} z_i^2},$$

where the z's are the least-squares residuals. The only difference from d is that one or two of the squared differences are omitted from the numerator of d'. Thus when m is small a substantial fraction of the relevant information is sacrificed.

The theory developed in Part I can be applied to show that d' lies between two values d'_L and d'_U. In contrast to d, exact significance points can be calculated for d'_L and d'_U using the results of R. L. Anderson (1942), except for the case of an even number of observations and an odd number of independent variables, for which the exact distribution of d'_L has not been found. A short table of such values for odd numbers of observations is given in Table 7. This table may be used for testing the significance of an observed value of d' in exactly the same way as Table 4 is used for testing the significance of an observed value of d.

Table 7. *Significance points of d'_L and d'_U*: 5 %

n	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	d'_L	d'_U	d'_L	d'_U	d'_L	d'_U	d'_L	d'_U	d'_L	d'_U
13	0·69	0·97	0·56	1·15	—	—	—	—	—	—
15	0·80	1·04	0·67	1·20	0·55	1·36	—	—	—	—
17	0·89	1·11	0·77	1·24	0·65	1·38	0·54	1·57	0·44	1·74
19	0·96	1·16	0·85	1·27	0·75	1·40	0·64	1·56	0·54	1·71
21	1·02	1·20	0·92	1·30	0·82	1·42	0·72	1·56	0·63	1·69
23	1·07	1·24	0·98	1·33	0·89	1·44	0·80	1·56	0·71	1·68

The calculations required even for such a short table were very heavy, and our chief motive for including it is the satisfaction of demonstrating that the problem has an exact solution. In practice we ourselves would prefer to use Table 4 owing to the greater power and simplicity of the statistic d.

We wish to express our most grateful thanks to Mrs E. G. Chambers and the computing staff of the Department of Applied Economics, Cambridge, for carying out most of the calculations required for Tables 4–7; also to Miss J. Graham of the Division of Research Techniques, London School of Economics, for assisting with Table 7.

REFERENCES

ANDERSON, R. L. (1942). *Ann. Math. Statist.* **13**, 1.
ANDERSON, R. L. & ANDERSON, T. W. (1950). *Ann. Math. Statist.* **21**, 59.
ANDERSON, R. L. & HOUSEMAN, E. E. (1942). Tables of orthogonal polynomial values extended to $N = 104$. *Res. Bull. Iowa St. Coll.* no. 297.
CARTER, A. H. (1947). *Biometrika*, **34**, 352.
COCHRAN, W. G. (1938). *J.R. Statist. Soc., Suppl.*, **5**, 171.
COURANT, R. & HILBERT, D. (1931). *Methoden der Mathematischen Physik*. Berlin: Julius Springer.
DURBIN, J. & WATSON, G. S. (1950). *Biometrika*, **37**, 409.
FISHER, R. A. & YATES, F. (1948). *Statistical Tables*. Edinburgh: Oliver and Boyd.
PEARSON, K. (1948). *Tables of the Incomplete Beta Function*. Cambridge University Press.
PREST, A. R. (1949). *Rev. Econ. Statist.* **31**, 33.
SCHULTZ, HENRY (1938). *Theory and Measurement of Demand*, pp. 674–7. University of Chicago Press.
SNEDECOR, G. W. (1937). *Statistical Methods*. Collegiate Press.
THOMPSON, CATHERINE (1941). *Biometrika*, **32**, 151.
WATSON, G. S. & DURBIN, J. (1951). Exact tests of serial correlation using non-circular statistics. (To be published).
WISE, M. E. (1950). *Biometrika*, **37**, 208.

CORRECTIONS TO PART I. (Durbin & Watson, 1950)

We are grateful to Prof. T. W. Anderson for pointing out an error in the section headed 'Inequalities on $\nu_1, \nu_2, \ldots, \nu_{n-k}$'. Part (*b*) of the lemma and its corollary have been correctly applied in the remaining parts of the paper, but are incorrectly stated.

The necessary corrections are as follows:

p. 415. The second paragraph beginning 'We therefore seek...' should read:

'We therefore seek inequalities on $\nu_1, \nu_2, \ldots, \nu_{n-k}$. For the sake of generality we suppose that certain of the regression vectors, say s of them, coincide with latent vectors of **A** (or are linear combinations of them). From the results of the previous section it follows that the problem is reduced to the consideration of $k-s$ arbitrary regression vectors, while **A** may be supposed to have s zero roots together with the roots of **A** not associated with the s latent vectors mentioned above. These roots may be renumbered so that

$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_{n-s}.$$

We proceed to show that $\lambda_i \leqslant \nu_i \leqslant \lambda_{i+k-s} \quad (i = 1, 2, \ldots, n-k).$' (3)

p. 415. The next to the last sentence should read:

'Allowing for cases in which regression vectors coincide with latent vectors of **A** we have (3).'

p. 416. Lemma (*b*) should read:

'If s of the columns of **X** are linear combinations of s of the latent vectors of **A**, and if the roots of **A** associated with the remaining $n-s$ latent vectors of **A** are renumbered so that

$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_{n-s},$$

then $\lambda_i \leqslant \nu_i \leqslant \lambda_{i+k-s} \quad (i = 1, 2, \ldots, n-k).$'

p. 416. Corollary should read: '$r_L \leqslant r \leqslant r_U$,

where $$r_L = \frac{\sum_{i=1}^{n-k} \lambda_i \zeta_i^2}{\sum_{i=1}^{n-k} \zeta_i^2}$$

and $$r_U = \frac{\sum_{i=1}^{n-k} \lambda_{i+k-s} \zeta_i^2}{\sum_{i=1}^{n-k} \zeta_i^2}.$$'

p. 416: line 16. 'λ_{s+k+1}' should read 'λ_{n-s+1}'.

The following misprints should also be noted.

p. 414, line 13. The equation should read:

$$\text{'}|\,\mathbf{I}_n\theta-(\mathbf{I}_n-\mathbf{l}_1\mathbf{l}_1')\,\mathbf{\Lambda}\,|=0.\text{'}$$

p. 414, line 22. The equation should read:

$$\text{'}\prod_{j=1}^{n}(\theta-\lambda_j)+\sum_{i=1}^{n}l_i^2\lambda_i\prod_{j\neq i}^{n}(\theta-\lambda_j)=0.\text{'}$$

p. 417, last line but two. The formula should read:

$$\text{'}\alpha_i=\prod_{j\neq i}^{\frac{1}{2}(n-k-1)}(\tau_i-\tau_j)\,\sqrt{(\tau_i-\tau)}.\text{'}$$

BI-VARIATE k-STATISTICS AND CUMULANTS OF THEIR JOINT SAMPLING DISTRIBUTION

By M. B. COOK, *University College, London*

1. The purpose of this paper is to derive

(a) formulae giving bivariate population moment-coefficients in terms of cumulants, and cumulants in terms of moment-coefficients up to the sixth order;

(b) formulae for the cumulants of the joint distribution of the simpler bivariate k-statistics.

In both cases these formulae are extensions, to two variates, of univariate formulae which have been fairly completely worked out in the past. But while the bivariate problem has been touched on in both aspects by various writers, only the simplest results have been published by way of illustration, and a need has been felt for a more extended list of formulae, partly on account of their intrinsic interest and partly to pave the way for further work in the direction of approximations to sampling distributions, along lines which have proved fruitful using combinations of univariate k-statistics.

2. Bivariate k-statistics are sample estimates of bivariate cumulants, and are expressible in terms of sample estimates of bivariate moment coefficients. It is desirable, therefore, to begin by giving the relations which exist between bivariate population moment-coefficients and cumulants. More than one method can be used to derive these.

(a) Using the generating functions we have

$$1+\frac{\mu'_{10}}{1!\,0!}t_1+\frac{\mu'_{01}}{0!\,1!}t_2+\ldots+\frac{\mu'_{pq}}{p!\,q!}t_1^p t_2^q+\ldots=\exp\left[\frac{\kappa_{10}}{1!\,0!}t_1+\frac{\kappa_{01}}{0!\,1!}t_2+\ldots+\frac{\kappa_{pq}}{p!\,q!}t_1^p t_2^q+\ldots\right]. \quad (1)$$

We may work out the coefficient of $t_1^p t_2^q/(p!\,q!)$ on the right side and equate it to the corresponding coefficient on the left, thus providing a formula for μ'_{pq} in terms of a series of cumulants. For example, the coefficient of $t_1 t_2^2$ on the right of (1) is

$$\frac{1}{1!}\left(\frac{\kappa_{12}}{1!\,2!}\right)+\frac{1}{2!}\left(\frac{\kappa_{02}}{0!\,2!}\frac{2\kappa_{10}}{1!\,0!}+\frac{\kappa_{11}}{1!\,1!}\frac{2\kappa_{01}}{0!\,1!}\right)+\frac{1}{3!}\left(3\frac{\kappa_{10}}{1!\,0!}\frac{\kappa_{01}}{0!\,1!}\frac{\kappa_{01}}{0!\,1!}\right),$$

so that

$$\mu'_{12}=\kappa_{12}+\kappa_{02}\kappa_{10}+2\kappa_{11}\kappa_{01}+\kappa_{01}^2\kappa_{10}.$$

Inversely, we can take the natural logarithm of both sides of (1) and work out the coefficient of $t_1^p t_2^q/(p!\,q!)$ on the left side, equating it to the corresponding coefficient on the right. This will give a formula for κ_{pq} in terms of a series of moment-coefficients about the origin.

By writing (1) as

$$1+\frac{\mu_{20}}{2!\,0!}t_1^2+\frac{\mu_{11}}{1!\,1!}t_1t_2+\frac{\mu_{02}}{0!\,2!}t_2^2+\ldots+\frac{\mu_{pq}}{p!\,q!}t_1^p t_2^q+\ldots$$
$$=\exp\left[\frac{\kappa_{20}}{2!\,0!}t_1^2+\frac{\kappa_{11}}{1!\,1!}t_1t_2+\frac{\kappa_{02}}{0!\,2!}t_2^2+\ldots+\frac{\kappa_{pq}}{p!\,q!}t_1^p t_2^q+\ldots\right], \quad (2)$$

in which the sum of the suffices of the μ's or κ's is at least 2, and proceeding similarly, we get a formula for μ_{pq}, the moment coefficient about the mean in terms of a series of cumulants

in which neither κ_{10} nor κ_{01} appears. The inverse formula is obtained by taking the natural logarithms of both sides.

(*b*) The process in (*a*) soon becomes laborious, but for any given order of result we can proceed, as indicated by Wishart (1928), by expanding (1) and (2) into their analogues for 3, 4 or more variates and then writing down a general result. Thus for three variates we should have

$$\mu'_{111} = \kappa_{111} + \kappa_{110}\kappa_{001} + \kappa_{101}\kappa_{010} + \kappa_{011}\kappa_{100} + \kappa_{100}\kappa_{010}\kappa_{001}, \tag{3}$$

which can, in fact, be derived directly by writing down all the combinations of one, two and three cumulants subject to the condition that the combined suffices should total 111. This is a simple combinatorial procedure. Suffices may then be amalgamated in (3), or extra noughts may be inserted. For example, adding the second and third suffices we have

$$\mu'_{12} = \kappa_{12} + 2\kappa_{11}\kappa_{01} + \kappa_{02}\kappa_{10} + \kappa_{10}\kappa_{01}^2.$$

Inversely we should note, for example, that

$$\kappa_{111} = \mu'_{111} - \mu'_{110}\mu'_{001} - \mu'_{101}\mu'_{010} - \mu'_{011}\mu'_{100} + 2\mu'_{100}\mu'_{010}\mu'_{001},$$

in which the term containing ρ moment-coefficients is multiplied by $(-1)^{\rho-1}(\rho-1)!$.

Corresponding to (2) we have an extension which, if taken to four variates, will give us

$$\mu'_{1111} = \kappa_{1111} + \kappa_{1100}\kappa_{0011} + \kappa_{1010}\kappa_{0101} + \kappa_{1001}\kappa_{0110},$$

which is, of course, the form assumed for μ'_{1111} when terms on the right containing unit parts are ignored. Inversely

$$\kappa_{1111} = \mu_{1111} - \mu_{1100}\mu_{0011} - \mu_{1010}\mu_{0101} - \mu_{1001}\mu_{0110}.$$

For example, amalgamating the first two and last two suffices we get

$$\mu_{22} = \kappa_{22} + \kappa_{20}\kappa_{02} + 2\kappa_{11}^2,$$
$$\kappa_{22} = \mu_{22} - \mu_{20}\mu_{02} - 2\mu_{11}^2.$$

Higher order results can easily be checked by noting from tables of symmetric functions (David & Kendall, 1949) that all partitions have been included and that the sum of terms of a particular partition is the partition coefficient given in the table.

(*c*) Kendall (1940) introduced a set of operators Δ_p, which, operating on the appropriate univariate formula, will give any desired bivariate formula. The technique, expressed in terms of an equivalent operator suggested by Kendall for another purpose, is best described by a specific example. We have

$$\mu'_5 = \kappa_5 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + 15\kappa_2^2\kappa_1 + 10\kappa_2\kappa_1^3 + \kappa_1^5$$

from published tables, or from the line $(1)^5$ in Table 1·5 of David & Kendall's tables of symmetric functions. Write this formally as

$$\mu'(r^5) = \kappa(r^5) + 5\kappa(r^4)\kappa(r) + 10\kappa(r^3)\kappa(r^2) + 10\kappa(r^3)\kappa^2(r) + 15\kappa^2(r^2)\kappa(r) + 10\kappa(r^2)\kappa^3(r) + \kappa^5(r),$$

and operate on it by $s\partial/\partial r$. Dividing by 5 we get

$$\begin{aligned}\mu'(r^4 s) = {} & \kappa(r^4 s) + \kappa(r^4)\kappa(s) + 4\kappa(r^3 s)\kappa(r) + 4\kappa(r^3)\kappa(rs) + 6\kappa(r^2 s)\kappa(r^2) + 4\kappa(r^3)\kappa(r)\kappa(s) \\ & + 6\kappa(r^2 s)\kappa^2(r) + 3\kappa^2(r^2)\kappa(s) + 12\kappa(r^2)\kappa(rs)\kappa(r) + 6\kappa(r^2)\kappa^2(r)\kappa(s) + 4\kappa(rs)\kappa^3(r) \\ & + \kappa^4(r)\kappa(s).\end{aligned}$$

Using the indices of r and s as suffices this gives

$$\mu'_{41} = \kappa_{41} + \kappa_{40}\kappa_{01} + 4\kappa_{31}\kappa_{10} + 4\kappa_{30}\kappa_{11} + 6\kappa_{21}\kappa_{20} + 4\kappa_{30}\kappa_{10}\kappa_{01} + 6\kappa_{21}\kappa_{10}^2 + 3\kappa_{20}^2\kappa_{01} + 12\kappa_{20}\kappa_{11}\kappa_{10} + 6\kappa_{20}\kappa_{10}^2\kappa_{01} + 4\kappa_{11}\kappa_{10}^3 + \kappa_{10}^4\kappa_{01}.$$

μ'_{32} may similarly be obtained by operating on $\mu'(r^4 s)$ by $s\partial/\partial r$. When we deal with moment-coefficients about the mean the operation is much simpler. We need only apply $s\partial/\partial r$ to

$$\mu(r^5) = \kappa(r^5) + 10\kappa(r^3)\kappa(r^2),$$

got from $\mu'(r^5)$ by deleting partitions having unit parts, to get

$$\mu_{41} = \kappa_{41} + 4\kappa_{30}\kappa_{11} + 6\kappa_{21}\kappa_{20}.$$

In both cases the inverse formulae can be written down at once.

Method (c) was in fact the one used. We give below all the formulae up to $p+q=6$, tabulating μ'_{pq} and κ_{pq} for $p \geqslant q$ only (interchange of suffices will give μ'_{qp} from μ'_{pq}), and incorporating for completeness the univariate formulae already published.

$$\mu'_{10} = \kappa_{10},$$

$$\mu'_{20} = \kappa_{20} + \kappa_{10}^2,$$

$$\mu'_{11} = \kappa_{11} + \kappa_{10}\kappa_{01},$$

$$\mu'_{30} = \kappa_{30} + 3\kappa_{20}\kappa_{10} + \kappa_{10}^3,$$

$$\mu'_{21} = \kappa_{21} + \kappa_{20}\kappa_{01} + 2\kappa_{11}\kappa_{10} + \kappa_{10}^2\kappa_{01},$$

$$\mu'_{40} = \kappa_{40} + 4\kappa_{30}\kappa_{10} + 3\kappa_{20}^2 + 6\kappa_{20}\kappa_{10}^2 + \kappa_{10}^4,$$

$$\mu'_{31} = \kappa_{31} + \kappa_{30}\kappa_{01} + 3\kappa_{21}\kappa_{10} + 3\kappa_{20}\kappa_{11} + 3\kappa_{20}\kappa_{10}\kappa_{01} + 3\kappa_{11}\kappa_{10}^2 + \kappa_{10}^3\kappa_{01},$$

$$\mu'_{22} = \kappa_{22} + 2\kappa_{21}\kappa_{01} + \kappa_{20}\kappa_{01}^2 + \kappa_{20}\kappa_{02} + 2\kappa_{12}\kappa_{10} + 2\kappa_{11}^2 + 4\kappa_{11}\kappa_{10}\kappa_{01} + \kappa_{10}^2\kappa_{01}^2 + \kappa_{10}^2\kappa_{02},$$

$$\mu'_{50} = \kappa_{50} + 5\kappa_{40}\kappa_{10} + 10\kappa_{30}\kappa_{20} + 10\kappa_{30}\kappa_{10}^2 + 15\kappa_{20}^2\kappa_{10} + 10\kappa_{20}\kappa_{10}^3 + \kappa_{10}^5,$$

$$\mu'_{41} = \kappa_{41} + \kappa_{40}\kappa_{01} + 4\kappa_{31}\kappa_{10} + 4\kappa_{30}\kappa_{11} + 4\kappa_{30}\kappa_{10}\kappa_{01} + 6\kappa_{21}\kappa_{20} + 6\kappa_{21}\kappa_{10}^2 + 3\kappa_{20}^2\kappa_{01} + 12\kappa_{20}\kappa_{11}\kappa_{10} + 6\kappa_{20}\kappa_{10}^2\kappa_{01} + 4\kappa_{11}\kappa_{10}^3 + \kappa_{10}^4\kappa_{01},$$

$$\mu'_{32} = \kappa_{32} + 2\kappa_{31}\kappa_{01} + \kappa_{30}\kappa_{01}^2 + \kappa_{30}\kappa_{02} + 3\kappa_{22}\kappa_{10} + 6\kappa_{21}\kappa_{11} + 6\kappa_{21}\kappa_{10}\kappa_{01} + 3\kappa_{20}\kappa_{12} + 6\kappa_{20}\kappa_{11}\kappa_{01} + 3\kappa_{20}\kappa_{10}\kappa_{01}^2 + 3\kappa_{20}\kappa_{10}\kappa_{02} + 3\kappa_{12}\kappa_{10}^2 + 6\kappa_{11}^2\kappa_{10} + 6\kappa_{11}\kappa_{10}^2\kappa_{01} + \kappa_{10}^3\kappa_{01}^2 + \kappa_{10}^3\kappa_{02},$$

$$\mu'_{60} = \kappa_{60} + 6\kappa_{50}\kappa_{10} + 15\kappa_{40}\kappa_{20} + 15\kappa_{40}\kappa_{10}^2 + 10\kappa_{30}^2 + 60\kappa_{30}\kappa_{20}\kappa_{10} + 20\kappa_{30}\kappa_{10}^3 + 15\kappa_{20}^3 + 45\kappa_{20}^2\kappa_{10}^2 + 15\kappa_{20}\kappa_{10}^4 + \kappa_{10}^6,$$

$$\mu'_{51} = \kappa_{51} + \kappa_{50}\kappa_{01} + 5\kappa_{41}\kappa_{10} + 5\kappa_{40}\kappa_{11} + 5\kappa_{40}\kappa_{10}\kappa_{01} + 10\kappa_{31}\kappa_{20} + 10\kappa_{31}\kappa_{10}^2 + 10\kappa_{30}\kappa_{21} + 10\kappa_{30}\kappa_{20}\kappa_{01} + 20\kappa_{30}\kappa_{11}\kappa_{10} + 10\kappa_{30}\kappa_{10}^2\kappa_{01} + 30\kappa_{21}\kappa_{20}\kappa_{10} + 10\kappa_{21}\kappa_{10}^3 + 15\kappa_{20}^2\kappa_{11} + 15\kappa_{20}^2\kappa_{10}\kappa_{01} + 30\kappa_{20}\kappa_{11}\kappa_{10}^2 + 10\kappa_{20}\kappa_{10}^3\kappa_{01} + 15\kappa_{11}\kappa_{10}^4 + \kappa_{10}^5\kappa_{01},$$

$$\mu'_{42} = \kappa_{42} + 2\kappa_{41}\kappa_{01} + \kappa_{40}\kappa_{01}^2 + \kappa_{40}\kappa_{02} + 4\kappa_{32}\kappa_{10} + 8\kappa_{31}\kappa_{11} + 8\kappa_{31}\kappa_{10}\kappa_{01} + 4\kappa_{30}\kappa_{12} + 8\kappa_{30}\kappa_{11}\kappa_{01} + 4\kappa_{30}\kappa_{10}\kappa_{01}^2 + 4\kappa_{30}\kappa_{10}\kappa_{02} + 6\kappa_{22}\kappa_{20} + 6\kappa_{22}\kappa_{10}^2 + 6\kappa_{21}^2 + 12\kappa_{21}\kappa_{20}\kappa_{01} + 24\kappa_{21}\kappa_{11}\kappa_{10} + 12\kappa_{21}\kappa_{10}^2\kappa_{01} + 3\kappa_{20}^2\kappa_{01}^2 + 3\kappa_{20}^2\kappa_{02} + 12\kappa_{20}\kappa_{12}\kappa_{10} + 12\kappa_{20}\kappa_{11}^2 + 24\kappa_{20}\kappa_{11}\kappa_{10}\kappa_{01} + 6\kappa_{20}\kappa_{10}^2\kappa_{01}^2 + 6\kappa_{20}\kappa_{10}^2\kappa_{02} + 4\kappa_{12}\kappa_{10}^3 + 12\kappa_{11}^2\kappa_{10}^2 + 8\kappa_{11}\kappa_{10}^3\kappa_{01} + \kappa_{10}^4\kappa_{01}^2 + \kappa_{10}^4\kappa_{02},$$

$$\mu'_{33} = \kappa_{33} + 3\kappa_{32}\kappa_{01} + 3\kappa_{31}\kappa_{01}^2 + 3\kappa_{31}\kappa_{02} + \kappa_{30}\kappa_{01}^3 + 3\kappa_{30}\kappa_{01}\kappa_{02} + \kappa_{30}\kappa_{03} + 3\kappa_{23}\kappa_{10} + 9\kappa_{22}\kappa_{11} + 9\kappa_{22}\kappa_{10}\kappa_{01} + 9\kappa_{21}\kappa_{12} + 18\kappa_{21}\kappa_{11}\kappa_{01} + 9\kappa_{21}\kappa_{10}\kappa_{01}^2 + 9\kappa_{21}\kappa_{10}\kappa_{02} + 3\kappa_{20}\kappa_{13} + 9\kappa_{20}\kappa_{12}\kappa_{01} + 9\kappa_{20}\kappa_{11}\kappa_{01}^2 + 9\kappa_{20}\kappa_{11}\kappa_{02} + 3\kappa_{20}\kappa_{10}\kappa_{01}^3 + 9\kappa_{20}\kappa_{10}\kappa_{01}\kappa_{02} + 3\kappa_{20}\kappa_{10}\kappa_{03} + 3\kappa_{13}\kappa_{10}^2 + 18\kappa_{12}\kappa_{11}\kappa_{10} + 9\kappa_{12}\kappa_{10}^2\kappa_{01} + 6\kappa_{11}^3 + 18\kappa_{11}^2\kappa_{10}\kappa_{01} + 9\kappa_{11}\kappa_{10}^2\kappa_{01}^2 + 9\kappa_{11}\kappa_{10}^2\kappa_{02} + \kappa_{10}^3\kappa_{01}^3 + 3\kappa_{10}^3\kappa_{01}\kappa_{02} + \kappa_{10}^3\kappa_{03}.$$

$$\kappa_{10} = \mu'_{10},$$

$$\kappa_{20} = \mu'_{20} - \mu'^{2}_{10},$$

$$\kappa_{11} = \mu'_{11} - \mu'_{10}\mu'_{01},$$

$$\kappa_{30} = \mu'_{30} - 3\mu'_{20}\mu'_{10} + 2\mu'^{3}_{10},$$

$$\kappa_{21} = \mu'_{21} - \mu'_{20}\mu'_{01} - 2\mu'_{11}\mu'_{10} + 2\mu'^{2}_{10}\mu'_{01},$$

$$\kappa_{40} = \mu'_{40} - 4\mu'_{30}\mu'_{10} - 3\mu'^{2}_{20} + 12\mu'_{20}\mu'^{2}_{10} - 6\mu'^{4}_{10},$$

$$\kappa_{31} = \mu'_{31} - \mu'_{30}\mu'_{01} - 3\mu'_{21}\mu'_{10} - 3\mu'_{20}\mu'_{11} + 6\mu'_{20}\mu'_{10}\mu'_{01} + 6\mu'_{11}\mu'^{2}_{10} - 6\mu'^{3}_{10}\mu'_{01},$$

$$\kappa_{22} = \mu'_{22} - 2\mu'_{21}\mu'_{01} + 2\mu'_{20}\mu'^{2}_{01} - \mu'_{20}\mu'_{02} - 2\mu'_{12}\mu'_{10} - 2\mu'^{2}_{11} + 8\mu'_{11}\mu'_{10}\mu'_{01} - 6\mu'^{2}_{10}\mu'^{2}_{01} + 2\mu'^{2}_{10}\mu'_{02},$$

$$\kappa_{50} = \mu'_{50} - 5\mu'_{40}\mu'_{10} - 10\mu'_{30}\mu'_{20} + 20\mu'_{30}\mu'^{2}_{10} + 30\mu'^{2}_{20}\mu'_{10} - 60\mu'_{20}\mu'^{3}_{10} + 24\mu'^{5}_{10},$$

$$\begin{aligned}\kappa_{41} = \mu'_{41} &- \mu'_{40}\mu'_{01} - 4\mu'_{31}\mu'_{10} - 4\mu'_{30}\mu'_{11} + 8\mu'_{30}\mu'_{10}\mu'_{01} - 6\mu'_{21}\mu'_{20} + 12\mu'_{21}\mu'^{2}_{10} + 6\mu'^{2}_{20}\mu'_{01} \\ &+ 24\mu'_{20}\mu'_{11}\mu'_{10} - 36\mu'_{20}\mu'^{2}_{10}\mu'_{01} - 24\mu'_{11}\mu'^{3}_{10} + 24\mu'^{4}_{10}\mu'_{01},\end{aligned}$$

$$\begin{aligned}\kappa_{32} = \mu'_{32} &- 2\mu'_{31}\mu'_{01} + 2\mu'_{30}\mu'^{2}_{01} - \mu'_{30}\mu'_{02} - 3\mu'_{22}\mu'_{10} - 6\mu'_{21}\mu'_{11} + 12\mu'_{21}\mu'_{10}\mu'_{01} - 3\mu'_{20}\mu'_{12} \\ &+ 12\mu'_{20}\mu'_{11}\mu'_{01} - 18\mu'_{20}\mu'_{10}\mu'^{2}_{01} + 6\mu'_{20}\mu'_{10}\mu'_{02} + 6\mu'_{12}\mu'^{2}_{10} + 12\mu'^{2}_{11}\mu'_{10} - 36\mu'_{11}\mu'^{2}_{10}\mu'_{01} \\ &+ 24\mu'^{3}_{10}\mu'^{2}_{01} - 6\mu'^{3}_{10}\mu'_{02},\end{aligned}$$

$$\begin{aligned}\kappa_{60} = \mu'_{60} &- 6\mu'_{50}\mu'_{10} - 15\mu'_{40}\mu'_{20} + 30\mu'_{40}\mu'^{2}_{10} - 10\mu'^{2}_{30} + 120\mu'_{30}\mu'_{20}\mu'_{10} - 120\mu'_{30}\mu'^{3}_{10} + 30\mu'^{3}_{20} \\ &- 270\mu'^{2}_{20}\mu'^{2}_{10} + 360\mu'_{20}\mu'^{4}_{10} - 120\mu'^{6}_{10},\end{aligned}$$

$$\begin{aligned}\kappa_{51} = \mu'_{51} &- \mu'_{50}\mu'_{01} - 5\mu'_{41}\mu'_{10} - 5\mu'_{40}\mu'_{11} + 10\mu'_{40}\mu'_{10}\mu'_{01} - 10\mu'_{31}\mu'_{20} + 20\mu'_{31}\mu'^{2}_{10} - 10\mu'_{30}\mu'_{21} \\ &+ 20\mu'_{30}\mu'_{20}\mu'_{01} + 40\mu'_{30}\mu'_{11}\mu'_{10} - 60\mu'_{30}\mu'^{2}_{10}\mu'_{01} + 60\mu'_{21}\mu'_{20}\mu'_{10} - 60\mu'_{21}\mu'^{3}_{10} + 30\mu'^{2}_{20}\mu'_{11} \\ &- 90\mu'^{2}_{20}\mu'_{10}\mu'_{01} - 180\mu'_{20}\mu'_{11}\mu'^{2}_{10} + 240\mu'_{20}\mu'^{3}_{10}\mu'_{01} + 120\mu'_{11}\mu'^{4}_{10} - 120\mu'^{5}_{10}\mu'_{01},\end{aligned}$$

$$\begin{aligned}\kappa_{42} = \mu'_{42} &- 2\mu'_{41}\mu'_{01} + 2\mu'_{40}\mu'^{2}_{01} - \mu'_{40}\mu'_{02} - 4\mu'_{32}\mu'_{10} - 8\mu'_{31}\mu'_{11} + 16\mu'_{31}\mu'_{10}\mu'_{01} - 4\mu'_{30}\mu'_{12} \\ &+ 16\mu'_{30}\mu'_{11}\mu'_{01} - 24\mu'_{30}\mu'_{10}\mu'^{2}_{01} + 8\mu'_{30}\mu'_{10}\mu'_{02} - 6\mu'_{22}\mu'_{20} + 12\mu'_{22}\mu'^{2}_{10} - 6\mu'^{2}_{21} \\ &+ 24\mu'_{21}\mu'_{20}\mu'_{01} + 48\mu'_{21}\mu'_{11}\mu'_{10} - 72\mu'_{21}\mu'^{2}_{10}\mu'_{01} - 18\mu'^{2}_{20}\mu'^{2}_{01} + 6\mu'^{2}_{20}\mu'_{02} + 24\mu'_{20}\mu'_{12}\mu'_{10} \\ &+ 24\mu'_{20}\mu'^{2}_{11} - 144\mu'_{20}\mu'_{11}\mu'_{10}\mu'_{01} + 144\mu'_{20}\mu'^{2}_{10}\mu'^{2}_{01} - 36\mu'_{20}\mu'^{2}_{10}\mu'_{02} - 24\mu'_{12}\mu'^{3}_{10} \\ &- 72\mu'^{2}_{11}\mu'^{2}_{10} + 192\mu'_{11}\mu'^{3}_{10}\mu'_{01} - 120\mu'^{4}_{10}\mu'^{2}_{01} + 24\mu'^{4}_{10}\mu'_{02},\end{aligned}$$

$$\begin{aligned}\kappa_{33} = \mu'_{33} &- 3\mu'_{32}\mu'_{01} + 6\mu'_{31}\mu'^{2}_{01} - 3\mu'_{31}\mu'_{02} - 6\mu'_{30}\mu'^{3}_{01} + 6\mu'_{30}\mu'_{01}\mu'_{02} - \mu'_{30}\mu'_{03} - 3\mu'_{23}\mu'_{10} \\ &- 9\mu'_{22}\mu'_{11} + 18\mu'_{22}\mu'_{10}\mu'_{01} - 9\mu'_{21}\mu'_{12} + 36\mu'_{21}\mu'_{11}\mu'_{01} - 54\mu'_{21}\mu'_{10}\mu'^{2}_{01} + 18\mu'_{21}\mu'_{10}\mu'_{02} \\ &- 3\mu'_{20}\mu'_{13} + 18\mu'_{20}\mu'_{12}\mu'_{01} - 54\mu'_{20}\mu'_{11}\mu'^{2}_{01} + 18\mu'_{20}\mu'_{11}\mu'_{02} + 72\mu'_{20}\mu'_{10}\mu'^{3}_{01} \\ &- 54\mu'_{20}\mu'_{10}\mu'_{01}\mu'_{02} + 6\mu'_{20}\mu'_{10}\mu'_{03} + 6\mu'_{13}\mu'^{3}_{10} + 36\mu'_{12}\mu'_{11}\mu'_{10} - 54\mu'_{12}\mu'^{2}_{10}\mu'_{01} + 12\mu'^{3}_{11} \\ &- 108\mu'^{2}_{11}\mu'_{10}\mu'_{01} + 216\mu'_{11}\mu'^{2}_{10}\mu'^{2}_{01} - 54\mu'_{11}\mu'^{2}_{10}\mu'_{02} - 120\mu'^{3}_{10}\mu'^{3}_{01} + 72\mu'^{3}_{10}\mu'_{01}\mu'_{02} \\ &- 6\mu'^{3}_{10}\mu'_{03}.\end{aligned}$$

When the origin is at the mean ($\kappa_{10} = \kappa_{01} = 0$):

$$\mu_{20} = \kappa_{20},$$

$$\mu_{11} = \kappa_{11},$$

$$\mu_{30} = \kappa_{30},$$

$$\mu_{21} = \kappa_{21},$$

$$\mu_{40} = \kappa_{40} + 3\kappa^{2}_{20},$$

$$\mu_{31} = \kappa_{31} + 3\kappa_{20}\kappa_{11},$$

$$\mu_{22} = \kappa_{22} + \kappa_{20}\kappa_{02} + 2\kappa^{2}_{11},$$

$$\mu_{50} = \kappa_{50} + 10\kappa_{30}\kappa_{20},$$

$$\mu_{41} = \kappa_{41} + 4\kappa_{30}\kappa_{11} + 6\kappa_{21}\kappa_{20},$$

$$\mu_{32} = \kappa_{32} + \kappa_{30}\kappa_{02} + 6\kappa_{21}\kappa_{11} + 3\kappa_{20}\kappa_{12},$$
$$\mu_{60} = \kappa_{60} + 15\kappa_{40}\kappa_{20} + 10\kappa_{30}^2 + 15\kappa_{20}^3,$$
$$\mu_{51} = \kappa_{51} + 5\kappa_{40}\kappa_{11} + 10\kappa_{31}\kappa_{20} + 10\kappa_{30}\kappa_{21} + 15\kappa_{20}^2\kappa_{11},$$
$$\mu_{42} = \kappa_{42} + \kappa_{40}\kappa_{02} + 8\kappa_{31}\kappa_{11} + 4\kappa_{30}\kappa_{12} + 6\kappa_{22}\kappa_{20} + 6\kappa_{21}^2 + 3\kappa_{20}^2\kappa_{02} + 12\kappa_{20}\kappa_{11}^2,$$
$$\mu_{33} = \kappa_{33} + 3\kappa_{31}\kappa_{02} + \kappa_{30}\kappa_{03} + 9\kappa_{22}\kappa_{11} + 9\kappa_{21}\kappa_{12} + 3\kappa_{20}\kappa_{13} + 9\kappa_{20}\kappa_{11}\kappa_{02} + 6\kappa_{11}^3,$$

and

$$\kappa_{20} = \mu_{20},$$
$$\kappa_{11} = \mu_{11},$$
$$\kappa_{30} = \mu_{30},$$
$$\kappa_{21} = \mu_{21},$$
$$\kappa_{40} = \mu_{40} - 3\mu_{20}^2,$$
$$\kappa_{31} = \mu_{31} - 3\mu_{20}\mu_{11},$$
$$\kappa_{22} = \mu_{22} - \mu_{20}\mu_{02} - 2\mu_{11}^2,$$
$$\kappa_{50} = \mu_{50} - 10\mu_{30}\mu_{20},$$
$$\kappa_{41} = \mu_{41} - 4\mu_{30}\mu_{11} - 6\mu_{21}\mu_{20},$$
$$\kappa_{32} = \mu_{32} - \mu_{30}\mu_{02} - 6\mu_{21}\mu_{11} - 3\mu_{20}\mu_{12},$$
$$\kappa_{60} = \mu_{60} - 15\mu_{40}\mu_{20} - 10\mu_{30}^2 + 30\mu_{20}^3,$$
$$\kappa_{51} = \mu_{51} - 5\mu_{40}\mu_{11} - 10\mu_{31}\mu_{20} - 10\mu_{30}\mu_{21} + 30\mu_{20}^2\mu_{11},$$
$$\kappa_{42} = \mu_{42} - \mu_{40}\mu_{02} - 8\mu_{31}\mu_{11} - 4\mu_{30}\mu_{12} - 6\mu_{22}\mu_{20} - 6\mu_{21}^2 + 6\mu_{20}^2\mu_{02} + 24\mu_{20}\mu_{11}^2,$$
$$\kappa_{33} = \mu_{33} - 3\mu_{31}\mu_{02} - \mu_{30}\mu_{03} - 9\mu_{22}\mu_{11} - 9\mu_{21}\mu_{12} - 3\mu_{20}\mu_{13} + 18\mu_{20}\mu_{11}\mu_{02} + 12\mu_{11}^3.$$

3. The univariate k-statistic k_r introduced by Fisher (1929) is a function of the sample observations $x_1, \ldots, x_n$ such that $\mathscr{E}(k_r) = \kappa_r$.

It may therefore be written as

$$k_r = \Sigma \frac{(-1)^{\rho-1}(\rho-1)!\,r!}{(r_1!)^{\pi_1}(r_2!)^{\pi_2}\ldots\pi_1!\ldots\pi_s!} \frac{[r_1^{\pi_1} r_2^{\pi_2}\ldots r_s^{\pi_s}]}{n^{(\rho)}},$$

in which the typical term corresponds to the partition $[r_1^{\pi_1} r_2^{\pi_2}\ldots]$ of the integer r, with $\sum_i (r_i\pi_i) = r$ and $\sum_i \pi_i = \rho$ and the summation is over all partitions. $[r_1^{\pi_1} r_2^{\pi_2}\ldots]$ is the augmented symmetric function of $x_1, x_2, \ldots, x_n$ (David & Kendall, 1949) and $n^{(\rho)}$ is the factorial expression $n(n-1)\ldots(n-\rho+1)$. To see that k_r has this form we need only note that the first part of the term under the summation sign is the coefficient of $\mu'^{\pi_1}_{r_1}\mu'^{\pi_2}_{r_2}\ldots$ in the expansion of κ_r in terms of moment coefficients about the origin, while

$$\mathscr{E}[r_1^{\pi_1} r_2^{\pi_2}\ldots] = n^{(\rho)}\mu'^{\pi_1}_{r_1}\mu'^{\pi_2}_{r_2}\ldots.$$

The tables of symmetric functions at once give k_r in terms of the power sums s_r, and they can be abbreviated by being expressed in terms of the sample estimates m_r of the moment coefficients about the mean (with the single exception that $k_1 = m'_1$, the sample mean).

In analogous fashion, if our sample consists of n pairs of observations (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) from a bivariate population, we define a k-statistic k_{pq} such that $\mathscr{E}(k_{pq}) = \kappa_{pq}$. It may then be written as

$$k_{pq} = \Sigma \frac{(-1)^{\rho-1}(\rho-1)!\,p!\,q!}{(p_1!\,q_1!)^{\pi_1}\pi_1!\,(p_2!\,q_2!)^{\pi_2}\pi_2!\ldots} \frac{[(p_1q_1)^{\pi_1}(p_2q_2)^{\pi_2}\ldots]}{n^{(\rho)}},$$

in which the unipartite number r is replaced by the bipartite number (pq) and the typical term corresponds to its partition $\{(p_1q_1)^{r_1}(p_2q_2)^{r_2}\ldots\}$. Values for k_{11}, k_{21}, k_{31} and k_{22} have been given by Fisher (1929) in terms of the power sums s_{pq} (k_{20}, k_{02}, etc., are univariate results). The tables we shall give only involve k_{11} and k_{21}, and these are very simply expressed as

$$k_{11} = \frac{n}{n-1} m_{11},$$

$$k_{21} = \frac{n^2}{(n-1)(n-2)} m_{21},$$

where m_{pq} is the sample moment coefficient, about the mean $(\bar{x}, \bar{y})$ corresponding to

$$m'_{pq} = \sum_{i=1}^{n} x_i^p y_i^q / n \text{ with } m'_{10} = \bar{x},\ m'_{01} = \bar{y}.$$

4. Our problem now resolves itself into the determination of the cumulants of the joint distribution of these bivariate k-statistics. To give a simple example, let $k_{\alpha\alpha'}$ and $k_{\beta\beta'}$ be two such statistics, and suppose we require the κ_{rs} of their joint distribution. To distinguish this from the population κ_{rs} we shall write it $\kappa[(\alpha\alpha')^r(\beta\beta')^s]$ with corresponding notations for the μ_{rs} and μ'_{rs} of the joint distribution. The more usual notation, which we preserve in the tables at the end of this paper, would be $\kappa\begin{pmatrix}\alpha \ldots \alpha & \beta \ldots \beta \\ \alpha' \ldots \alpha' & \beta \ldots \beta'\end{pmatrix}$, but for simplicity in writing and setting the present notation is better. The first step in the algebraic derivation is always to calculate

$$\mu'[(\alpha\alpha')^r(\beta\beta')^s] = \mathscr{E}(k_{\alpha\alpha'}^r k_{\beta\beta'}^s),$$

which will be given in terms of population bivariate moment-coefficients about the mean if partitions involving unit parts are ignored. These can be converted into population cumulants. We then invoke the formulae of § 2 which, in the present notation, will give $\mu'[(\alpha\alpha')^r(\beta\beta')^s]$ in terms of $\kappa[(\alpha\alpha')^r(\beta\beta')^s]$ and other terms which are combinations of lower order κ's, which will have been calculated already. By removing all surplus terms from $\mu'[(\alpha\alpha')^r(\beta\beta')^s]$ we are finally left with $\kappa[(\alpha\alpha')^r(\beta\beta')^s]$. Up to the third order the result is identical with that for

$$\mu[(\alpha\alpha')^r(\beta\beta')^s] = \mathscr{E}\{(k_{\alpha\alpha'} - \kappa_{\alpha\alpha'})^r (k_{\beta\beta'} - \kappa_{\beta\beta'})^s\}, \quad (r+s \leqslant 3).$$

Purely algebraic derivation of even the comparatively simple results given at the end of this paper would be a formidable task, and we seek other methods. One which naturally commends itself is Fisher's combinatorial technique (Fisher, 1929) which is applied throughout as in the case of the univariate k-statistics, but distinguishing two kinds of objects in the various partitions. The n-coefficients, which depend only on the patterns, remain the same, and we have the advantage of a check in the sense that a single term of a univariate k-statistic result has its numerical coefficient already known, and we can be reasonably sure of our accuracy if the coefficients of the subpartitions add up to this total. To illustrate, let us find the coefficient of $k_{32}k_{20}^2$ in $\kappa[(30)(21)^2]$. The corresponding univariate result is the coefficient of $k_5 k_2^2$ in $\kappa(3^3)$. For this the partitions are

(*a*)

3	1	1	5
0	1	1	2
0	1	1	2
3	3	3	

(*b*)

2	2	1	5
1	0	1	2
0	1	1	2
3	3	3	

These contribute $54/\{(n-1)(n-2)\}$ and $162/(n-1)^2$ respectively. To deal with the bivariate problem we must alter the column entries so that in two columns the three 1's are replaced by two 1's and an X. We then find that (a) evolves into six partitions, two of which have row totals (32), (20), (20). They are

3	X	X	(32)
0	1	1	(20)
0	1	1	(20)
(30)	(21)	(21)	

$(2X)$	1	X	(32)
0	1	1	(20)
0	1	1	(20)
(21)	(30)	(21)	

The numerical coefficients, i.e. the numbers of ways of assigning the objects subject to the row conditions, are 2 and 12 respectively. This gives a contribution $14/\{(n-1)(n-2)\}$ to the desired result. On the other hand, (b) evolves into nine partitions, of which two concern us, namely,

2	$(1X)$	X	(32)
1	0	1	(20)
0	1	1	(20)
(30)	(21)	(21)	

$(1X)$	$(1X)$	1	(32)
1	0	1	(20)
0	1	1	(20)
(21)	(21)	(30)	

the numerical coefficients are both 24, thus providing a contribution $48/(n-1)^2$. Finally, the coefficient of $k_{32}k_{20}^2$ in $\kappa[(30)(21)^2]$ is

$$\frac{14}{(n-1)(n-2)}+\frac{48}{(n-1)^2}=\frac{2(31n-55)}{(n-1)^2(n-2)}.$$

The other partitions provide the coefficients of $k_{50}k_{20}k_{02}$, $k_{50}k_{11}^2$ and $k_{41}k_{20}k_{11}$. The reader who cares to complete the process will verify that (a) yields

$$\frac{2}{(n-1)(n-2)}(2\kappa_{50}\kappa_{20}\kappa_{02}+2\kappa_{50}\kappa_{11}^2+16\kappa_{41}\kappa_{20}\kappa_{11}+7\kappa_{32}\kappa_{20}^2),$$

while (b) similarly yields

$$\frac{6}{(n-1)^2}(2\kappa_{50}\kappa_{20}\kappa_{02}+3\kappa_{50}\kappa_{11}^2+14\kappa_{41}\kappa_{20}\kappa_{11}+8\kappa_{32}\kappa_{20}^2).$$

Adding term by term gives us four terms of $\kappa[(30)(21)^2]$.

5. In the particular case of the joint sampling distribution of k_{20}, k_{02} and k_{11} in samples from a normal bivariate population, Wishart (1928) showed how it was possible to generalize the results. He also generalized the procedure to deal with as many variates (beyond two) as might be required, and gave the completely general cumulants, to the fourth order, of the simultaneous distribution of the second-order moment statistics, on the assumption of normality of the parent multivariate population. All special cases can be derived from these general results by coalescing the variates in all possible ways.

6. The method we have followed in deriving the results listed at the end of this paper, which are for samples from a general (infinite) population, is the operational method due to Kendall (1940), essentially an extension of the method already used in § 2. We shall illustrate by working out all the terms of $\kappa[(30)(21)^2]$ which correspond to that of $k_5k_2^2$ in $\kappa[3^3]$. Write $\kappa[3^3]\equiv\kappa[(30)^3]$ formally as $\kappa[r_1^3r_2^3r_3^3]$ and operate on it by

$$S=s_1s_2\frac{\partial^2}{\partial r_1\partial r_2}+s_2s_3\frac{\partial^2}{\partial r_2\partial r_3}+s_3s_1\frac{\partial^2}{\partial r_3\partial r_1},$$

obtaining

$$9\kappa[r_1^2s_1r_2^2s_2r_3^3]+9\kappa[r_1^3r_2^2s_2r_3^2s_3]+9[r_1^2s_1r_2^3r_3^2s_3].$$

Retaining the indices only of the r's and s's this becomes $27\kappa[(30)(21)^2]$, since the order in which the constituents k_{30}, k_{21} and k_{12} are written is immaterial. We have therefore to operate with $S/27$ on expressions which correspond with the partitions (a) and (b) of § 4. The first, taking account of row composition, is written as $\kappa(r_1^3 r_2 r_3)\,\kappa^2(r_2 r_3)$ and operating with $S/27$ yields

$$\begin{aligned}\tfrac{1}{27}[&\kappa(r_1^3 s_2 s_3)\,\kappa^2(r_2 r_3) + 2\kappa(r_1^3 s_2 r_3)\,\kappa(r_2 r_3)\,\kappa(r_2 s_3) + 2\kappa(r_1^3 r_2 r_3)\,\kappa(r_2 r_3)\,\kappa(s_2 s_3)\\ &+ 2\kappa(r_1^3 r_2 r_3)\,\kappa(r_2 s_3)\,\kappa(s_2 r_3) + 2\kappa(r_1^3 r_2 s_3)\,\kappa(r_2 r_3)\,\kappa(s_2 r_3) + 3\kappa(r_1^2 s_1 s_2 r_3)\,\kappa^2(r_2 r_3)\\ &+ 6\kappa(r_1^2 s_1 r_2 r_3)\,\kappa(r_2 r_3)\,\kappa(s_2 r_3) + 3\kappa(r_1^2 s_1 r_2 s_3)\,\kappa^2(r_2 r_3) + 6\kappa(r_1^2 s_1 r_2 r_3)\,\kappa(r_2 r_3)\,\kappa(r_2 s_3)].\end{aligned}$$

Using indices only of the r's and s's, and collecting terms, this reduces to

$$\tfrac{1}{27}[2\kappa_{50}\kappa_{20}\kappa_{02} + 2\kappa_{50}\kappa_{11}^2 + 16\kappa_{41}\kappa_{20}\kappa_{11} + 7\kappa_{32}\kappa_{20}^2].$$

Noting that the coefficient of partition (a) of § 4 is $54/\{(n-1)(n-2)\}$ we see that this partition contributes the term

$$\frac{2}{(n-1)(n-2)}[2\kappa_{50}\kappa_{20}\kappa_{02} + 2\kappa_{50}\kappa_{11}^2 + 16\kappa_{41}\kappa_{20}\kappa_{11} + 7\kappa_{32}\kappa_{20}^2]$$

to the desired result. This agrees with the result partially worked out in § 4 by the combinatorial method.

The expression corresponding to partition (b) of § 4, which has coefficient $162/(n-1)^2$, is written $\kappa(r_1^2 r_2^2 r_3)\,\kappa(r_1 r_3)\,\kappa(r_2 r_3)$, and operating similarly with $S/27$ yields finally the terms

$$\frac{6}{(n-1)^2}[2\kappa_{50}\kappa_{20}\kappa_{02} + 3\kappa_{50}\kappa_{11}^2 + 14\kappa_{41}\kappa_{20}\kappa_{11} + 8\kappa_{32}\kappa_{20}^2].$$

Adding the two expressions term by term we have the final result that the terms of $\kappa[(30)(21)^2]$ corresponding to that of $\kappa_5\kappa_2^2$ in $\kappa[3^3]$ are

$$\begin{aligned}\frac{2}{(n-1)^2(n-2)}\{&2(4n-7)\,\kappa_{50}\kappa_{20}\kappa_{02} + (11n-20)\,\kappa_{50}\kappa_{11}^2\\ &+ 2(29n-50)\,\kappa_{41}\kappa_{20}\kappa_{11} + (31n-55)\,\kappa_{32}\kappa_{20}^2\}.\end{aligned}$$

By using this operator $S/27$ on *all* the partitions which contribute to the terms of $k[3^3]$ we build up a complete expression for $\kappa[(30)(21)^2]$. For every such third-order cumulant, based on $\kappa[3^3]$, of the joint distribution, there is an appropriate operator. For example, $\kappa[(21)^2(12)]$ may be obtained by using the operator

$$\frac{1}{162}\Sigma s_1 s_2^2 s_3 \frac{\partial^4}{\partial r_1 \partial r_2^2 \partial r_3}.$$

Similarly, cumulants of *all* orders of the joint distribution may be worked out by choosing the operator appropriately.

7. The procedure indicated in § 6 is somewhat laborious, but is straightforward and therefore avoids errors of omission. It is not, however, necessary to obtain every single result by a separate operation. Consider, for example, the univariate result

$$\kappa[2^3] = \frac{\kappa_6}{n^2} + \frac{12\kappa_4\kappa_2}{n(n-1)} + \frac{4(n-2)}{n(n-1)^2}\kappa_3^2 + \frac{8}{(n-1)^2}\kappa_2^4,$$

corresponding to the partitions

2	2	2	6
2	2	2	

2	1	1	4
0	1	1	2
2	2	2	

1	1	1	3
1	1	1	3
2	2	2	

1	1	0	2
1	0	1	2
0	1	1	2
2	2	2	

Formally the result may be written

$$\kappa[r_1^2 r_2^2 r_3^2] = \frac{1}{n^2}\kappa(r_1^2 r_2^2 r_3^2) + \frac{12}{n(n-1)}\kappa(r_1^2 r_2 r_3)\,\kappa(r_2 r_3) + \frac{4(n-2)}{n(n-1)^2}\kappa^2(r_1 r_2 r_3) + \frac{8}{(n-1)^2}\kappa(r_1 r_2)\,\kappa(r_1 r_3)\,\kappa(r_2 r_3).$$

Operation by $\Sigma s_i \dfrac{\partial}{\partial r_i}$ $(i = 1, 2, 3)$ will yield $\kappa[(20)^2(11)]$. But this is evidently equivalent to the easier process of operating by $s\partial/\partial r$ on

$$\kappa[(r^2)^3] = \frac{1}{n^2}\kappa(r^6) + \frac{12}{n(n-1)}\kappa(r^4)\,\kappa(r^2) + \frac{4(n-2)}{n(n-1)^2}\kappa^2(r^3) + \frac{8}{(n-1)^2}\kappa^3(r^2).$$

Similarly,
$$\sum_{i=1}^{3} s_i^2 \frac{\partial^2}{\partial r_i^2} + \sum_{i \neq j}\sum s_i s_j \frac{\partial^2}{\partial r_i \partial r_j} \equiv s^2 \frac{\partial^2}{\partial r^2},$$

so that operation by $s^2\partial^2/\partial r^2$ will produce $\kappa[(20)^2(02)] + 2\kappa[(20)(11)^2]$. If one of these results has been obtained by the procedure of § 6 the other may be obtained by difference if the relatively easy operation by $s^2\partial^2/\partial r^2$ is carried out. In general, it is found that for any given order approximately half the cumulants of the joint distribution must be worked out by the methods of § 4 or § 6, while the remainder can be obtained by using the appropriate power of $s\partial/\partial r$ as the operator.

As already indicated, the method used to obtain the formulae which follow was that of § 6, but the operator discussed in the present section has provided a useful check on the results. Since most of the results are new it was considered desirable to work them out by two different methods.

$$\kappa\begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix} = \frac{1}{n}\kappa_{40} + \frac{2}{n-1}\kappa_{20}^2,$$

$$\kappa\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} = \frac{1}{n}\kappa_{31} + \frac{2}{n-1}\kappa_{20}\kappa_{11},$$

$$\kappa\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{2}{n-1}\kappa_{11}^2,$$

$$\kappa\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{1}{n-1}\kappa_{20}\kappa_{02} + \frac{1}{n-1}\kappa_{11}^2.$$

$$\kappa\begin{pmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \end{pmatrix} = \frac{1}{n^2}\kappa_{60} + \frac{12}{n(n-1)}\kappa_{40}\kappa_{20} + \frac{4(n-2)}{n(n-1)^2}\kappa_{30}^2 + \frac{8}{(n-1)^2}\kappa_{20}^3,$$

$$\kappa\begin{pmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \frac{1}{n^2}\kappa_{51} + \frac{8}{n(n-1)}\kappa_{31}\kappa_{20} + \frac{4}{n(n-1)}\kappa_{40}\kappa_{11} + \frac{4(n-2)}{n(n-1)^2}\kappa_{30}\kappa_{21} + \frac{8}{(n-1)^2}\kappa_{20}^2\kappa_{11},$$

$$\kappa\begin{pmatrix} 2 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \frac{1}{n^2}\kappa_{42} + \frac{4}{n(n-1)}\kappa_{22}\kappa_{20} + \frac{8}{n(n-1)}\kappa_{31}\kappa_{11} + \frac{4(n-2)}{n(n-1)^2}\kappa_{21}^2 + \frac{8}{(n-1)^2}\kappa_{20}\kappa_{11}^2,$$

$$\kappa\begin{pmatrix}2&1&1\\0&1&1\end{pmatrix}=\frac{1}{n^2}\kappa_{42}+\frac{5}{n(n-1)}\kappa_{22}\kappa_{20}+\frac{6}{n(n-1)}\kappa_{31}\kappa_{11}+\frac{1}{n(n-1)}\kappa_{40}\kappa_{02}+\frac{2(n-2)}{n(n-1)^2}\kappa_{21}^2$$
$$+\frac{2(n-2)}{n(n-1)^2}\kappa_{30}\kappa_{12}+\frac{6}{(n-1)^2}\kappa_{20}\kappa_{11}^2+\frac{2}{(n-1)^2}\kappa_{20}^2\kappa_{02},$$

$$\kappa\begin{pmatrix}2&0&1\\0&2&1\end{pmatrix}=\frac{1}{n^2}\kappa_{33}+\frac{8}{n(n-1)}\kappa_{22}\kappa_{11}+\frac{2}{n(n-1)}\kappa_{20}\kappa_{13}+\frac{2}{n(n-1)}\kappa_{31}\kappa_{02}+\frac{4(n-2)}{n(n-1)^2}\kappa_{21}\kappa_{12}$$
$$+\frac{4}{(n-1)^2}\kappa_{11}^3+\frac{4}{(n-1)^2}\kappa_{11}\kappa_{20}\kappa_{02},$$

$$\kappa\begin{pmatrix}1&1&1\\1&1&1\end{pmatrix}=\frac{1}{n^2}\kappa_{33}+\frac{6}{n(n-1)}\kappa_{22}\kappa_{11}+\frac{3}{n(n-1)}\kappa_{20}\kappa_{13}+\frac{3}{n(n-1)}\kappa_{31}\kappa_{02}+\frac{3(n-2)}{n(n-1)^2}\kappa_{21}\kappa_{12}$$
$$+\frac{(n-2)}{n(n-1)^2}\kappa_{30}\kappa_{03}+\frac{2}{(n-1)^2}\kappa_{11}^3+\frac{6}{(n-1)^2}\kappa_{11}\kappa_{20}\kappa_{02}.$$

$$\kappa\begin{pmatrix}2&2&2&2\\0&0&0&0\end{pmatrix}=\frac{1}{n^3}\kappa_{80}+\frac{24}{n^2(n-1)}\kappa_{60}\kappa_{20}+\frac{32(n-2)}{n^2(n-1)^2}\kappa_{50}\kappa_{30}+\frac{8(4n^2-9n+6)}{n^2(n-1)^3}\kappa_{40}^2$$
$$+\frac{144}{n(n-1)^2}\kappa_{40}\kappa_{20}^2+\frac{96(n-2)}{n(n-1)^3}\kappa_{30}^2\kappa_{20}+\frac{48}{(n-1)^3}\kappa_{20}^4,$$

$$\kappa\begin{pmatrix}2&2&2&1\\0&0&0&1\end{pmatrix}=\frac{1}{n^3}\kappa_{71}+\frac{18}{n^2(n-1)}\kappa_{51}\kappa_{20}+\frac{6}{n^2(n-1)}\kappa_{60}\kappa_{11}+\frac{20(n-2)}{n^2(n-1)^2}\kappa_{41}\kappa_{30}+\frac{12(n-2)}{n^2(n-1)^2}\kappa_{50}\kappa_{21}$$
$$+\frac{8(4n^2-9n+6)}{n^2(n-1)^3}\kappa_{40}\kappa_{31}+\frac{72}{n(n-1)^2}\kappa_{40}\kappa_{11}\kappa_{20}+\frac{72}{n(n-1)^2}\kappa_{31}\kappa_{20}^2+\frac{24(n-2)}{n(n-1)^3}\kappa_{30}^2\kappa_{11}$$
$$+\frac{72(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{20}\kappa_{21}+\frac{48}{(n-1)^3}\kappa_{20}^3\kappa_{11},$$

$$\kappa\begin{pmatrix}2&2&2&0\\0&0&0&2\end{pmatrix}=\frac{1}{n^3}\kappa_{62}+\frac{12}{n^2(n-1)}\kappa_{51}\kappa_{11}+\frac{12}{n^2(n-1)}\kappa_{42}\kappa_{20}+\frac{24(n-2)}{n^2(n-1)^2}\kappa_{41}\kappa_{21}+\frac{8(n-2)}{n^2(n-1)^2}\kappa_{32}\kappa_{30}$$
$$+\frac{4(5n^2-12n+9)}{n^2(n-1)^3}\kappa_{31}^2+\frac{12}{n^2(n-1)}\kappa_{40}\kappa_{22}+\frac{24}{n(n-1)^2}\kappa_{40}\kappa_{11}^2+\frac{24}{n(n-1)^2}\kappa_{22}\kappa_{20}^2$$
$$+\frac{96}{n(n-1)^2}\kappa_{31}\kappa_{20}\kappa_{11}+\frac{48(n-2)}{n(n-1)^3}\kappa_{21}^2\kappa_{20}+\frac{48(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{21}\kappa_{11}+\frac{48}{(n-1)^3}\kappa_{20}^2\kappa_{11}^2,$$

$$\kappa\begin{pmatrix}2&2&1&1\\0&0&1&1\end{pmatrix}=\frac{1}{n^3}\kappa_{62}+\frac{13}{n^2(n-1)}\kappa_{42}\kappa_{20}+\frac{10}{n^2(n-1)}\kappa_{51}\kappa_{11}+\frac{1}{n^2(n-1)}\kappa_{60}\kappa_{02}+\frac{12(n-2)}{n^2(n-1)^2}\kappa_{32}\kappa_{30}$$
$$+\frac{16(n-2)}{n^2(n-1)^2}\kappa_{41}\kappa_{21}+\frac{4(n-2)}{n^2(n-1)^2}\kappa_{50}\kappa_{12}+\frac{2(7n^2-16n+11)}{n^2(n-1)^3}\kappa_{40}\kappa_{22}$$
$$+\frac{2(9n^2-20n+13)}{n^2(n-1)^3}\kappa_{31}^2+\frac{80}{n(n-1)^2}\kappa_{31}\kappa_{20}\kappa_{11}+\frac{12}{n(n-1)^2}\kappa_{40}\kappa_{20}\kappa_{02}$$
$$+\frac{32}{n(n-1)^2}\kappa_{22}\kappa_{20}^2+\frac{20}{n(n-1)^2}\kappa_{40}\kappa_{11}^2+\frac{40(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{21}\kappa_{11}+\frac{4(n-2)}{n(n-1)^3}\kappa_{30}^2\kappa_{02}$$
$$+\frac{28(n-2)}{n(n-1)^3}\kappa_{21}^2\kappa_{20}+\frac{24(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{20}\kappa_{12}+\frac{40}{(n-1)^3}\kappa_{20}^2\kappa_{11}^2+\frac{8}{(n-1)^3}\kappa_{20}^3\kappa_{02},$$

$$\kappa\begin{pmatrix}2&2&1&0\\0&0&1&2\end{pmatrix}=\frac{1}{n^3}\kappa_{53}+\frac{14}{n^2(n-1)}\kappa_{42}\kappa_{11}+\frac{8}{n^2(n-1)}\kappa_{33}\kappa_{20}+\frac{2}{n^2(n-1)}\kappa_{51}\kappa_{02}+\frac{20(n-2)}{n^2(n-1)^2}\kappa_{32}\kappa_{21}$$
$$+\frac{8(n-2)}{n^2(n-1)^2}\kappa_{41}\kappa_{12}+\frac{4(n-2)}{n^2(n-1)^2}\kappa_{30}\kappa_{23}+\frac{4(7n^2-16n+11)}{n^2(n-1)^3}\kappa_{31}\kappa_{22}+\frac{4}{n^2(n-1)}\kappa_{40}\kappa_{13}$$
$$+\frac{48}{n(n-1)^2}\kappa_{31}\kappa_{11}^2+\frac{64}{n(n-1)^2}\kappa_{11}\kappa_{22}\kappa_{20}+\frac{8}{n(n-1)^2}\kappa_{20}^2\kappa_{13}+\frac{8}{n(n-1)^2}\kappa_{40}\kappa_{11}\kappa_{02}$$
$$+\frac{16}{n(n-1)^2}\kappa_{31}\kappa_{20}\kappa_{02}+\frac{40(n-2)}{n(n-1)^3}\kappa_{21}^2\kappa_{11}+\frac{8(n-2)}{n(n-1)^3}\kappa_{21}\kappa_{30}\kappa_{02}$$
$$+\frac{32(n-2)}{n(n-1)^3}\kappa_{20}\kappa_{21}\kappa_{12}+\frac{16(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{11}\kappa_{12}+\frac{32}{(n-1)^3}\kappa_{11}^3\kappa_{20}+\frac{16}{(n-1)^3}\kappa_{11}\kappa_{20}^2\kappa_{02},$$

$$\kappa\begin{pmatrix}2&1&1&1\\0&1&1&1\end{pmatrix}=\frac{1}{n^3}\kappa_{53}+\frac{9}{n^2(n-1)}\kappa_{33}\kappa_{20}+\frac{12}{n^2(n-1)}\kappa_{42}\kappa_{11}+\frac{3}{n^2(n-1)}\kappa_{51}\kappa_{02}+\frac{7(n-2)}{n^2(n-1)^2}\kappa_{23}\kappa_{30}$$
$$+\frac{15(n-2)}{n^2(n-1)^2}\kappa_{32}\kappa_{21}+\frac{9(n-2)}{n^2(n-1)^2}\kappa_{41}\kappa_{12}+\frac{(n-2)}{n^2(n-1)^2}\kappa_{50}\kappa_{03}+\frac{5n^2-12n+9}{n^2(n-1)^3}\kappa_{40}\kappa_{13}$$
$$+\frac{3(9n^2-20n+13)}{n^2(n-1)^3}\kappa_{31}\kappa_{22}+\frac{36}{n(n-1)^2}\kappa_{31}\kappa_{11}^2+\frac{60}{n(n-1)^2}\kappa_{11}\kappa_{20}\kappa_{22}$$
$$+\frac{24}{n(n-1)^2}\kappa_{31}\kappa_{20}\kappa_{02}+\frac{12}{n(n-1)^2}\kappa_{40}\kappa_{11}\kappa_{02}+\frac{12}{n(n-1)^2}\kappa_{20}^2\kappa_{13}+\frac{24(n-2)}{n(n-1)^3}\kappa_{21}^2\kappa_{11}$$
$$+\frac{12(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{21}\kappa_{02}+\frac{30(n-2)}{n(n-1)^3}\kappa_{20}\kappa_{21}\kappa_{12}+\frac{24(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{12}\kappa_{11}$$
$$+\frac{6(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{03}\kappa_{20}+\frac{24}{(n-1)^3}\kappa_{20}\kappa_{11}^3+\frac{24}{(n-1)^3}\kappa_{20}^2\kappa_{02}\kappa_{11},$$

$$\kappa\begin{pmatrix}2&2&0&0\\0&0&2&2\end{pmatrix}=\frac{1}{n^3}\kappa_{44}+\frac{16}{n^2(n-1)}\kappa_{33}\kappa_{11}+\frac{4}{n^2(n-1)}\kappa_{24}\kappa_{20}+\frac{4}{n^2(n-1)}\kappa_{42}\kappa_{02}+\frac{16(n-2)}{n^2(n-1)^2}\kappa_{32}\kappa_{12}$$
$$+\frac{16(n-2)}{n^2(n-1)^2}\kappa_{23}\kappa_{21}+\frac{8(2n^2-5n+4)}{n^2(n-1)^3}\kappa_{22}^2+\frac{16}{n^2(n-1)}\kappa_{31}\kappa_{13}+\frac{64}{n(n-1)^2}\kappa_{11}^2\kappa_{22}$$
$$+\frac{32}{n(n-1)^2}\kappa_{20}\kappa_{11}\kappa_{13}+\frac{32}{n(n-1)^2}\kappa_{31}\kappa_{11}\kappa_{02}+\frac{16}{n(n-1)^2}\kappa_{22}\kappa_{20}\kappa_{02}+\frac{16(n-2)}{n(n-1)^3}\kappa_{21}^2\kappa_{02}$$
$$+\frac{16(n-2)}{n(n-1)^3}\kappa_{12}^2\kappa_{20}+\frac{64(n-2)}{n(n-1)^3}\kappa_{11}\kappa_{21}\kappa_{12}+\frac{32}{(n-1)^3}\kappa_{11}^3\kappa_{20}\kappa_{02}+\frac{16}{(n-1)^3}\kappa_{11}^4,$$

$$\kappa\begin{pmatrix}2&0&1&1\\0&2&1&1\end{pmatrix}=\frac{1}{n^3}\kappa_{44}+\frac{5}{n^2(n-1)}\kappa_{24}\kappa_{20}+\frac{14}{n^2(n-1)}\kappa_{33}\kappa_{11}+\frac{5}{n^2(n-1)}\kappa_{42}\kappa_{02}+\frac{14(n-2)}{n^2(n-1)^2}\kappa_{23}\kappa_{21}$$
$$+\frac{14(n-2)}{n^2(n-1)^2}\kappa_{32}\kappa_{12}+\frac{2(n-2)}{n^2(n-1)^2}\kappa_{41}\kappa_{03}+\frac{2(n-2)}{n^2(n-1)^2}\kappa_{14}\kappa_{30}$$
$$+\frac{2(7n^2-16n+11)}{n^2(n-1)^3}\kappa_{31}\kappa_{13}+\frac{(17n^2-38n+25)}{n^2(n-1)^3}\kappa_{22}^2+\frac{1}{n^2(n-1)}\kappa_{40}\kappa_{04}$$
$$+\frac{52}{n(n-1)^2}\kappa_{22}\kappa_{11}^2+\frac{24}{n(n-1)^2}\kappa_{22}\kappa_{20}\kappa_{02}+\frac{32}{n(n-1)^2}\kappa_{11}\kappa_{20}\kappa_{13}+\frac{32}{n(n-1)^2}\kappa_{11}\kappa_{02}\kappa_{31}$$
$$+\frac{2}{n(n-1)^2}\kappa_{40}\kappa_{02}^2+\frac{2}{n(n-1)^2}\kappa_{20}^2\kappa_{04}+\frac{12(n-2)}{n(n-1)^3}\kappa_{21}^2\kappa_{02}+\frac{52(n-2)}{n(n-1)^3}\kappa_{11}\kappa_{21}\kappa_{12}$$
$$+\frac{12(n-2)}{n(n-1)^3}\kappa_{20}\kappa_{12}^2+\frac{8(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{12}\kappa_{02}+\frac{4(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{03}\kappa_{11}+\frac{8(n-2)}{n(n-1)^3}\kappa_{21}\kappa_{03}\kappa_{20}$$
$$+\frac{12}{(n-1)^3}\kappa_{11}^4+\frac{32}{(n-1)^3}\kappa_{11}^2\kappa_{20}\kappa_{02}+\frac{4}{(n-1)^3}\kappa_{20}^2\kappa_{02}^2,$$

$$\kappa\begin{pmatrix}1&1&1&1\\1&1&1&1\end{pmatrix}=\frac{1}{n^3}\kappa_{44}+\frac{6}{n^2(n-1)}\kappa_{24}\kappa_{20}+\frac{12}{n^2(n-1)}\kappa_{33}\kappa_{11}+\frac{6}{n^2(n-1)}\kappa_{42}\kappa_{02}+\frac{4(n-2)}{n^2(n-1)^2}\kappa_{14}\kappa_{30}$$
$$+\frac{12(n-2)}{n^2(n-1)^2}\kappa_{23}\kappa_{21}+\frac{12(n-2)}{n^2(n-1)^2}\kappa_{32}\kappa_{12}+\frac{4(n-2)}{n^2(n-1)^2}\kappa_{41}\kappa_{03}+\frac{(n^2-3n+3)}{n^2(n-1)^3}\kappa_{40}\kappa_{04}$$
$$+\frac{3(5n^2-11n+7)}{n^2(n-1)^3}\kappa_{22}^2+\frac{4(4n^2-3n+6)}{n^2(n-1)^3}\kappa_{31}\kappa_{13}+\frac{30}{n(n-1)^2}\kappa_{20}\kappa_{02}\kappa_{22}$$
$$+\frac{36}{n(n-1)^2}\kappa_{11}\kappa_{13}\kappa_{20}+\frac{36}{n(n-1)^2}\kappa_{11}\kappa_{02}\kappa_{31}+\frac{36}{n(n-1)^2}\kappa_{22}\kappa_{11}^2+\frac{3}{n(n-1)^2}\kappa_{20}^2\kappa_{04}$$
$$+\frac{3}{n(n-1)^2}\kappa_{02}^2\kappa_{40}+\frac{36(n-2)}{n(n-1)^3}\kappa_{21}\kappa_{12}\kappa_{11}+\frac{12(n-2)}{n(n-1)^3}\kappa_{11}\kappa_{30}\kappa_{03}+\frac{12(n-2)}{n(n-1)^3}\kappa_{21}^2\kappa_{02}$$
$$+\frac{12(n-2)}{n(n-1)^3}\kappa_{30}\kappa_{12}\kappa_{02}+\frac{12(n-2)}{n(n-1)^3}\kappa_{12}^2\kappa_{20}+\frac{12(n-2)}{n(n-1)^3}\kappa_{20}\kappa_{21}\kappa_{03}+\frac{36}{(n-1)^3}\kappa_{11}^2\kappa_{20}\kappa_{02}$$
$$+\frac{6}{(n-1)^3}\kappa_{20}^2\kappa_{02}^2+\frac{6}{(n-1)^3}\kappa_{11}^4.$$

$$\kappa\begin{pmatrix}3&3\\0&0\end{pmatrix}=\frac{1}{n}\kappa_{60}+\frac{9}{n-1}\kappa_{40}\kappa_{20}+\frac{9}{n-1}\kappa_{30}^2+\frac{6n}{(n-1)(n-2)}\kappa_{20}^3,$$

$$\kappa\begin{pmatrix}3&2\\0&1\end{pmatrix}=\frac{1}{n}\kappa_{51}+\frac{6}{n-1}\kappa_{31}\kappa_{20}+\frac{3}{n-1}\kappa_{40}\kappa_{11}+\frac{9}{n-1}\kappa_{30}\kappa_{21}+\frac{6n}{(n-1)(n-2)}\kappa_{20}^2\kappa_{11},$$

$$\kappa\begin{pmatrix}3&1\\0&2\end{pmatrix}=\frac{1}{n}\kappa_{42}+\frac{3}{n-1}\kappa_{22}\kappa_{20}+\frac{6}{n-1}\kappa_{31}\kappa_{11}+\frac{3}{n-1}\kappa_{30}\kappa_{12}+\frac{6}{n-1}\kappa_{21}^2+\frac{6n}{(n-1)(n-2)}\kappa_{20}\kappa_{11}^2,$$

$$\kappa\begin{pmatrix}2&2\\1&1\end{pmatrix}=\frac{1}{n}\kappa_{42}+\frac{4}{n-1}\kappa_{22}\kappa_{20}+\frac{4}{n-1}\kappa_{31}\kappa_{11}+\frac{1}{n-1}\kappa_{40}\kappa_{02}+\frac{4}{n-1}\kappa_{30}\kappa_{12}+\frac{5}{n-1}\kappa_{21}^2$$
$$+\frac{4n}{(n-1)(n-2)}\kappa_{20}\kappa_{11}^2+\frac{2n}{(n-1)(n-2)}\kappa_{20}^2\kappa_{02},$$

$$\kappa\begin{pmatrix}3&0\\0&3\end{pmatrix}=\frac{1}{n}\kappa_{33}+\frac{9}{n-1}\kappa_{22}\kappa_{11}+\frac{9}{n-1}\kappa_{12}\kappa_{21}+\frac{6n}{(n-1)(n-2)}\kappa_{11}^3,$$

$$\kappa\begin{pmatrix}2&1\\1&2\end{pmatrix}=\frac{1}{n}\kappa_{33}+\frac{5}{n-1}\kappa_{22}\kappa_{11}+\frac{2}{n-1}\kappa_{13}\kappa_{20}+\frac{2}{n-1}\kappa_{31}\kappa_{02}+\frac{1}{n-1}\kappa_{30}\kappa_{03}+\frac{8}{n-1}\kappa_{21}\kappa_{12}$$
$$+\frac{4n}{(n-1)(n-2)}\kappa_{11}\kappa_{20}\kappa_{02}+\frac{2n}{(n-1)(n-2)}\kappa_{11}^3.$$

$$\kappa\begin{pmatrix}3&3&3\\0&0&0\end{pmatrix}=\frac{1}{n^2}\kappa_{90}+\frac{27}{n(n-1)}\kappa_{70}\kappa_{20}+\frac{27(3n-4)}{n(n-1)^2}\kappa_{60}\kappa_{30}+\frac{27(4n-7)}{n(n-1)^2}\kappa_{50}\kappa_{40}$$
$$+\frac{54(4n-7)}{(n-1)^2(n-2)}\kappa_{50}\kappa_{20}^2+\frac{162(5n-12)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{30}\kappa_{20}+\frac{36(7n^2-30n+34)}{(n-1)^2(n-2)^2}\kappa_{30}^3$$
$$+\frac{108n(5n-12)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{20}^3,$$

$$\kappa\begin{pmatrix}3&3&2\\0&0&1\end{pmatrix}=\frac{1}{n^2}\kappa_{81}+\frac{21}{n(n-1)}\kappa_{61}\kappa_{20}+\frac{6}{n(n-1)}\kappa_{70}\kappa_{11}+\frac{9(3n-4)}{n(n-1)^2}\kappa_{60}\kappa_{21}+\frac{18(3n-4)}{n(n-1)^2}\kappa_{51}\kappa_{30}$$
$$+\frac{12(4n-7)}{n(n-1)^2}\kappa_{50}\kappa_{31}+\frac{15(4n-7)}{n(n-1)^2}\kappa_{41}\kappa_{40}+\frac{24(4n-7)}{(n-1)^2(n-2)}\kappa_{50}\kappa_{20}\kappa_{11}$$
$$+\frac{30(4n-7)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{20}^2+\frac{72(5n-12)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{30}\kappa_{20}+\frac{54(5n-12)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{21}\kappa_{20}$$
$$+\frac{36(5n-12)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{30}\kappa_{11}+\frac{36(7n^2-30n+34)}{(n-1)^2(n-2)^2}\kappa_{30}^2\kappa_{21}+\frac{36n(5n-12)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{20}^3$$
$$+\frac{72n(5n-12)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{20}^2\kappa_{11},$$

$$\kappa\begin{pmatrix}3&3&1\\0&0&2\end{pmatrix}=\frac{1}{n^2}\kappa_{72}+\frac{12}{n(n-1)}\kappa_{61}\kappa_{11}+\frac{15}{n(n-1)}\kappa_{52}\kappa_{20}+\frac{6}{n(n-1)}\kappa_{60}\kappa_{12}+\frac{6(7n-10)}{n(n-1)^2}\kappa_{51}\kappa_{21}$$
$$+\frac{3(11n-14)}{n(n-1)^2}\kappa_{42}\kappa_{30}+\frac{3(5n-8)}{n(n-1)^2}\kappa_{50}\kappa_{22}+\frac{6(11n-20)}{n(n-1)^2}\kappa_{41}\kappa_{31}+\frac{9(3n-5)}{n(n-1)^2}\kappa_{32}\kappa_{40}$$
$$+\frac{6(5n-8)}{(n-1)^2(n-2)}\kappa_{50}\kappa_{11}^2+\frac{12(11n-20)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{11}\kappa_{20}+\frac{18(3n-5)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{20}^2$$
$$+\frac{72(3n-7)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{30}\kappa_{11}+\frac{144(2n-5)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{21}\kappa_{20}+\frac{36(3n-7)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{30}\kappa_{20}$$
$$+\frac{72(2n-5)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{21}\kappa_{11}+\frac{54}{(n-1)^2}\kappa_{40}\kappa_{12}\kappa_{20}+\frac{54}{(n-1)^2}\kappa_{30}^2\kappa_{12}$$
$$+\frac{18(11n^2-48n+56)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{21}^2+\frac{36n}{(n-1)^2(n-2)}\kappa_{12}\kappa_{20}^3+\frac{144n(2n-5)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{20}^2\kappa_{11}$$
$$+\frac{72n(3n-7)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{20}\kappa_{11}^2,$$

$$\kappa\begin{pmatrix}3&2&2\\0&1&1\end{pmatrix}=\frac{1}{n^2}\kappa_{72}+\frac{1}{n(n-1)}\kappa_{70}\kappa_{02}+\frac{10}{n(n-1)}\kappa_{61}\kappa_{11}+\frac{16}{n(n-1)}\kappa_{52}\kappa_{20}+\frac{(7n-10)}{n(n-1)^2}\kappa_{60}\kappa_{12}$$
$$+\frac{4(10n-13)}{n(n-1)^2}\kappa_{51}\kappa_{21}+\frac{2(17n-23)}{n(n-1)^2}\kappa_{42}\kappa_{30}+\frac{(19n-34)}{n(n-1)^2}\kappa_{50}\kappa_{22}$$
$$+\frac{2(29n-50)}{n(n-1)^2}\kappa_{41}\kappa_{31}+\frac{(31n-55)}{n(n-1)^2}\kappa_{32}\kappa_{40}+\frac{2(11n-20)}{(n-1)^2(n-2)}\kappa_{50}\kappa_{11}^2$$
$$+\frac{4(4n-7)}{(n-1)^2(n-2)}\kappa_{50}\kappa_{20}\kappa_{02}+\frac{4(29n-50)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{11}\kappa_{20}+\frac{2(31n-55)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{20}^2$$
$$+\frac{24(7n-17)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{30}\kappa_{11}+\frac{24(11n-26)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{21}\kappa_{20}+\frac{12(12n-29)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{30}\kappa_{20}$$
$$+\frac{12(11n-26)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{21}\kappa_{11}+\frac{36(2n-5)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{12}\kappa_{20}+\frac{6(5n-12)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{30}\kappa_{02}$$
$$+\frac{6(11n^2-48n+56)}{(n-1)^2(n-2)^2}\kappa_{30}^2\kappa_{12}+\frac{6(31n^2-132n+148)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{21}^2$$
$$+\frac{24n(2n-5)}{(n-1)^2(n-2)^2}\kappa_{12}\kappa_{20}^3+\frac{24n(11n-26)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{20}^2\kappa_{11}+\frac{12n(5n-12)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{20}^2\kappa_{02}$$
$$+\frac{24n(7n-17)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{20}\kappa_{11}^2,$$

$$\kappa\begin{pmatrix}2&2&2\\1&1&1\end{pmatrix} = \frac{1}{n^2}\kappa_{63} + \frac{3}{n(n-1)}\kappa_{61}\kappa_{02} + \frac{12}{n(n-1)}\kappa_{52}\kappa_{11} + \frac{12}{n(n-1)}\kappa_{43}\kappa_{20} + \frac{6(3n-4)}{n(n-1)^2}\kappa_{51}\kappa_{12}$$
$$+ \frac{6(7n-9)}{n(n-1)^2}\kappa_{42}\kappa_{21} + \frac{4(5n-7)}{n(n-1)^2}\kappa_{33}\kappa_{30} + \frac{(n-2)}{n(n-1)^2}\kappa_{60}\kappa_{03} + \frac{6(n-2)}{n(n-1)^2}\kappa_{50}\kappa_{13}$$
$$+ \frac{3(13n-22)}{n(n-1)^2}\kappa_{41}\kappa_{22} + \frac{12(4n-7)}{n(n-1)^2}\kappa_{32}\kappa_{31} + \frac{3(5n-9)}{n(n-1)^2}\kappa_{23}\kappa_{40} + \frac{12}{(n-1)^2}\kappa_{50}\kappa_{11}\kappa_{02}$$
$$+ \frac{6(7n-12)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{11}^2 + \frac{12(3n-5)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{20}\kappa_{02} + \frac{24(4n-7)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{20}\kappa_{11}$$
$$+ \frac{6(5n-9)}{(n-1)^2(n-2)}\kappa_{23}\kappa_{20}^2 + \frac{24(2n-5)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{30}\kappa_{02} + \frac{24(8n-19)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{21}\kappa_{11}$$
$$+ \frac{12(9n-22)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{30}\kappa_{11} + \frac{24(5n-12)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{12}\kappa_{20} + \frac{60(3n-7)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{21}\kappa_{20}$$
$$+ \frac{24(2n-5)}{(n-1)^2(n-2)}\kappa_{13}\kappa_{30}\kappa_{20} + \frac{6(7n-16)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{21}\kappa_{02} + \frac{12(5n-12)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{12}\kappa_{11}$$
$$+ \frac{12(n-3)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{03}\kappa_{20} + \frac{24(7n^2-30n+34)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{21}\kappa_{12}$$
$$+ \frac{2(5n^2-24n+32)}{(n-1)^2(n-2)^2}\kappa_{30}^2\kappa_{03} + \frac{2(37n^2-156n+172)}{(n-1)^2(n-2)^2}\kappa_{21}^3 + \frac{8n(n-3)}{(n-1)^2(n-2)^2}\kappa_{03}\kappa_{20}^3$$
$$+ \frac{12n(7n-16)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{20}^2\kappa_{02} + \frac{24n(8n-19)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{20}\kappa_{11}^2$$
$$+ \frac{24n(5n-12)}{(n-1)^2(n-2)^2}\kappa_{12}\kappa_{20}^2\kappa_{11} + \frac{48n(2n-5)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{20}\kappa_{11}\kappa_{02}$$
$$+ \frac{8n(5n-12)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{11}^3,$$

$$\kappa\begin{pmatrix}3&3&0\\0&0&3\end{pmatrix} = \frac{1}{n^2}\kappa_{63} + \frac{18}{n(n-1)}\kappa_{52}\kappa_{11} + \frac{9}{n(n-1)}\kappa_{43}\kappa_{20} + \frac{18}{n(n-1)}\kappa_{51}\kappa_{12} + \frac{9(5n-8)}{n(n-1)^2}\kappa_{42}\kappa_{21}$$
$$+ \frac{18}{n(n-1)}\kappa_{33}\kappa_{30} + \frac{9}{n(n-1)}\kappa_{23}\kappa_{40} + \frac{9(5n-8)}{n(n-1)^2}\kappa_{41}\kappa_{22} + \frac{54(n-2)}{n(n-1)^2}\kappa_{32}\kappa_{31}$$
$$+ \frac{18(5n-8)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{11}^2 + \frac{108}{(n-1)^2}\kappa_{32}\kappa_{20}\kappa_{11} + \frac{18}{(n-1)(n-2)}\kappa_{23}\kappa_{20}^2$$
$$+ \frac{108(3n-8)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{21}\kappa_{11} + \frac{162}{(n-1)^2}\kappa_{22}\kappa_{30}\kappa_{11} + \frac{108}{(n-1)^2}\kappa_{31}\kappa_{12}\kappa_{20}$$
$$+ \frac{54(3n-8)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{21}\kappa_{20} + \frac{54}{(n-1)^2}\kappa_{40}\kappa_{12}\kappa_{11} + \frac{162}{(n-1)^2}\kappa_{30}\kappa_{21}\kappa_{12}$$
$$+ \frac{18(5n^2-24n+32)}{(n-1)^2(n-2)^2}\kappa_{21}^3 + \frac{108n(3n-8)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{20}\kappa_{11}^2 + \frac{108n}{(n-1)^2(n-2)}\kappa_{12}\kappa_{20}^2\kappa_{11}$$
$$+ \frac{108n}{(n-1)^2(n-2)}\kappa_{30}\kappa_{11}^3,$$

$$\kappa\begin{pmatrix}3&2&1\\0&1&2\end{pmatrix}=\frac{1}{n^2}\kappa_{63}+\frac{11}{n(n-1)}\kappa_{43}\kappa_{20}+\frac{14}{n(n-1)}\kappa_{52}\kappa_{11}+\frac{2}{n(n-1)}\kappa_{61}\kappa_{02}+\frac{1}{n(n-1)}\kappa_{60}\kappa_{03}$$
$$+\frac{(17n-23)}{n(n-1)^2}\kappa_{51}\kappa_{12}+\frac{(44n-59)}{n(n-1)^2}\kappa_{42}\kappa_{21}+\frac{(19n-25)}{n(n-1)^2}\kappa_{33}\kappa_{30}+\frac{(5n-8)}{n(n-1)^2}\kappa_{50}\kappa_{13}$$
$$+\frac{2(19n-34)}{n(n-1)^2}\kappa_{41}\kappa_{22}+\frac{(53n-92)}{n(n-1)^2}\kappa_{32}\kappa_{31}+\frac{3(4n-7)}{n(n-1)^2}\kappa_{23}\kappa_{40}$$
$$+\frac{2(5n-8)}{(n-1)^2(n-2)}\kappa_{50}\kappa_{11}\kappa_{02}+\frac{6(9n-16)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{11}^2+\frac{2(11n-20)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{20}\kappa_{02}$$
$$+\frac{2(53n-92)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{20}\kappa_{11}+\frac{6(4n-7)}{(n-1)^2(n-2)}\kappa_{23}\kappa_{20}^2+\frac{12(3n-7)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{30}\kappa_{02}$$
$$+\frac{6(41n-98)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{21}\kappa_{11}+\frac{3(39n-94)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{30}\kappa_{11}+\frac{6(19n-46)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{12}\kappa_{20}$$
$$+\frac{9(19n-46)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{21}\kappa_{20}+\frac{12(3n-7)}{(n-1)^2(n-2)}\kappa_{13}\kappa_{30}\kappa_{20}+\frac{12(2n-5)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{21}\kappa_{02}$$
$$+\frac{3(19n-46)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{12}\kappa_{11}+\frac{9}{(n-1)^2}\kappa_{40}\kappa_{03}\kappa_{20}+\frac{3(53n^2-228n+260)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{21}\kappa_{12}$$
$$+\frac{9}{(n-1)^2}\kappa_{30}^2\kappa_{03}+\frac{12(7n^2-30n+34)}{(n-1)^2(n-2)^2}\kappa_{21}^3+\frac{6n}{(n-1)^2(n-2)}\kappa_{03}\kappa_{20}^3$$
$$+\frac{24n(2n-5)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{20}^2\kappa_{02}+\frac{6n(41n-98)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{20}\kappa_{11}^2$$
$$+\frac{6n(19n-46)}{(n-1)^2(n-2)^2}\kappa_{12}\kappa_{20}^2\kappa_{11}+\frac{24n(3n-7)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{20}\kappa_{11}\kappa_{02}$$
$$+\frac{6n(9n-22)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{11}^3,$$

$$\kappa\begin{pmatrix}2&2&1\\1&1&2\end{pmatrix}=\frac{1}{n^2}\kappa_{54}+\frac{14}{n(n-1)}\kappa_{43}\kappa_{11}+\frac{5}{n(n-1)}\kappa_{52}\kappa_{02}+\frac{8}{n(n-1)}\kappa_{34}\kappa_{20}+\frac{(29n-38)}{n(n-1)^2}\kappa_{42}\kappa_{12}$$
$$+\frac{2(2n-3)}{n(n-1)^2}\kappa_{51}\kappa_{03}+\frac{2(19n-25)}{n(n-1)^2}\kappa_{33}\kappa_{21}+\frac{2(5n-7)}{n(n-1)^2}\kappa_{24}\kappa_{30}+\frac{(n-2)}{n(n-1)^2}\kappa_{50}\kappa_{04}$$
$$+\frac{2(9n-16)}{n(n-1)^2}\kappa_{41}\kappa_{13}+\frac{2(25n-43)}{n(n-1)^2}\kappa_{32}\kappa_{22}+\frac{2(17n-30)}{n(n-1)^2}\kappa_{23}\kappa_{31}+\frac{(5n-9)}{n(n-1)^2}\kappa_{14}\kappa_{40}$$
$$+\frac{2}{(n-1)^2}\kappa_{50}\kappa_{02}^2+\frac{4(9n-16)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{11}\kappa_{02}+\frac{4(15n-26)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{11}^2$$
$$+\frac{4(10n-17)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{20}\kappa_{02}+\frac{4(17n-30)}{(n-1)^2(n-2)}\kappa_{23}\kappa_{20}\kappa_{11}+\frac{2(5n-9)}{(n-1)^2(n-2)}\kappa_{14}\kappa_{20}^2$$
$$+\frac{4(21n-50)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{21}\kappa_{02}+\frac{2(21n-52)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{30}\kappa_{02}+\frac{8(18n-43)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{12}\kappa_{11}$$
$$+\frac{10(21n-50)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{21}\kappa_{11}+\frac{4(13n-32)}{(n-1)^2(n-2)}\kappa_{13}\kappa_{30}\kappa_{11}+\frac{4(7n-18)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{03}\kappa_{20}$$
$$+\frac{4(30n-71)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{12}\kappa_{20}+\frac{4(21n-50)}{(n-1)^2(n-2)}\kappa_{13}\kappa_{21}\kappa_{20}+\frac{4(2n-5)}{(n-1)^2(n-2)}\kappa_{30}\kappa_{20}\kappa_{04}$$
$$+\frac{8(3n-7)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{12}\kappa_{02}+\frac{2(7n-18)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{03}\kappa_{11}$$

$$+\frac{2(19n^2-84n+100)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{03}\kappa_{21}+\frac{2(79n^2-336n+376)}{(n-1)^2(n-2)^2}\kappa_{21}^2\kappa_{12}$$

$$+\frac{8(7n^2-30n+34)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{12}^2+\frac{4n(7n-18)}{(n-1)^2(n-2)^2}\kappa_{20}^2\kappa_{11}\kappa_{03}$$

$$+\frac{8n(21n-50)}{(n-1)^2(n-2)^2}\kappa_{20}\kappa_{21}\kappa_{11}\kappa_{02}+\frac{16n(3n-7)}{(n-1)^2(n-2)^2}\kappa_{20}^2\kappa_{02}\kappa_{12}$$

$$+\frac{4n(21n-50)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{11}^3+\frac{8n(18n-43)}{(n-1)^2(n-2)^2}\kappa_{11}^2\kappa_{20}\kappa_{12}+\frac{8n(2n-5)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{20}\kappa_{02}^2$$

$$+\frac{4n(13n-32)}{(n-1)(n-2)^2}\kappa_{11}^2\kappa_{30}\kappa_{02},$$

$$\kappa\begin{pmatrix}3&1&1\\0&2&2\end{pmatrix}=\frac{1}{n^2}\kappa_{54}+\frac{16}{n(n-1)}\kappa_{43}\kappa_{11}+\frac{4}{n(n-1)}\kappa_{52}\kappa_{02}+\frac{7}{n(n-1)}\kappa_{34}\kappa_{20}+\frac{4}{n(n-1)}\kappa_{51}\kappa_{03}$$

$$+\frac{4(7n-10)}{n(n-1)^2}\kappa_{42}\kappa_{12}+\frac{4(10n-13)}{n(n-1)^2}\kappa_{33}\kappa_{21}+\frac{3(3n-4)}{n(n-1)^2}\kappa_{24}\kappa_{30}+\frac{1}{n(n-1)}\kappa_{50}\kappa_{04}$$

$$+\frac{4(4n-7)}{n(n-1)^2}\kappa_{41}\kappa_{13}+\frac{2(26n-47)}{n(n-1)^2}\kappa_{32}\kappa_{22}+\frac{12(3n-5)}{n(n-1)^2}\kappa_{23}\kappa_{31}+\frac{3(n-2)}{n(n-1)^2}\kappa_{14}\kappa_{40}$$

$$+\frac{2}{(n-1)(n-2)}\kappa_{50}\kappa_{02}^2+\frac{8(4n-7)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{11}\kappa_{02}+\frac{4(19n-34)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{11}^2$$

$$+\frac{4(7n-13)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{20}\kappa_{02}+\frac{24(3n-5)}{(n-1)^2(n-2)}\kappa_{23}\kappa_{20}\kappa_{11}+\frac{6}{(n-1)^2}\kappa_{14}\kappa_{20}^2$$

$$+\frac{24(3n-7)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{21}\kappa_{02}+\frac{12(3n-7)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{30}\kappa_{02}+\frac{24(7n-17)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{12}\kappa_{11}$$

$$+\frac{12(21n-50)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{21}\kappa_{11}+\frac{24(2n-5)}{(n-1)^2(n-2)}\kappa_{13}\kappa_{30}\kappa_{11}+\frac{24}{(n-1)^2}\kappa_{31}\kappa_{03}\kappa_{20}$$

$$+\frac{12(9n-23)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{12}\kappa_{20}+\frac{24(3n-7)}{(n-1)^2(n-2)}\kappa_{13}\kappa_{21}\kappa_{20}+\frac{6}{(n-1)^2}\kappa_{04}\kappa_{30}\kappa_{20}$$

$$+\frac{12(n-3)}{(n-1)^2(n-2)}\kappa_{40}\kappa_{12}\kappa_{02}+\frac{12}{(n-1)^2}\kappa_{40}\kappa_{03}\kappa_{11}+\frac{36}{(n-1)^2}\kappa_{30}\kappa_{03}\kappa_{21}$$

$$+\frac{24(7n^2-30n+34)}{(n-1)^2(n-2)^2}\kappa_{21}^2\kappa_{12}+\frac{24(2n^2-9n+11)}{(n-1)^2(n-2)^2}\kappa_{12}^2\kappa_{30}+\frac{24n}{(n-1)^2(n-2)}\kappa_{20}^2\kappa_{11}\kappa_{03}$$

$$+\frac{48n(3n-7)}{(n-1)^2(n-2)^2}\kappa_{20}\kappa_{21}\kappa_{11}\kappa_{02}+\frac{24n(n-3)}{(n-1)^2(n-2)^2}\kappa_{12}\kappa_{20}^2\kappa_{02}$$

$$+\frac{24n(5n-12)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{11}^3+\frac{24n(7n-17)}{(n-1)^2(n-2)^2}\kappa_{12}\kappa_{20}\kappa_{11}^2+\frac{12n}{(n-1)^2(n-2)}\kappa_{30}\kappa_{20}\kappa_{02}^2$$

$$+\frac{24n(2n-5)}{(n-1)^2(n-2)^2}\kappa_{30}\kappa_{11}^2\kappa_{02},$$

$$\kappa\begin{pmatrix}3&0&2\\0&3&1\end{pmatrix}=\frac{1}{n^2}\kappa_{54}+\frac{18}{n(n-1)}\kappa_{43}\kappa_{11}+\frac{3}{n(n-1)}\kappa_{52}\kappa_{02}+\frac{6}{n(n-1)}\kappa_{34}\kappa_{20}+\frac{3}{n(n-1)}\kappa_{51}\kappa_{03}$$

$$+\frac{3(10n-13)}{n(n-1)^2}\kappa_{42}\kappa_{12}+\frac{3(13n-19)}{n(n-1)^2}\kappa_{33}\kappa_{21}+\frac{9}{n(n-1)}\kappa_{24}\kappa_{30}+\frac{3(5n-8)}{n(n-1)^2}\kappa_{41}\kappa_{13}$$

$$+\frac{3(19n-34)}{n(n-1)^2}\kappa_{32}\kappa_{22}+\frac{3(11n-20)}{n(n-1)^2}\kappa_{23}\kappa_{31}+\frac{3}{n(n-1)}\kappa_{14}\kappa_{40}$$
$$+\frac{6(5n-8)}{(n-1)^2(n-2)}\kappa_{41}\kappa_{11}\kappa_{02}+\frac{24(4n-7)}{(n-1)^2(n-2)}\kappa_{32}\kappa_{11}^2+\frac{18}{(n-1)^2}\kappa_{32}\kappa_{20}\kappa_{02}$$
$$+\frac{6(11n-20)}{(n-1)^2(n-2)}\kappa_{23}\kappa_{20}\kappa_{11}+\frac{6}{(n-1)(n-2)}\kappa_{14}\kappa_{20}^2+\frac{18(3n-8)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{21}\kappa_{02}$$
$$+\frac{27}{(n-1)^2}\kappa_{22}\kappa_{30}\kappa_{02}+\frac{36(5n-12)}{(n-1)^2(n-2)}\kappa_{31}\kappa_{12}\kappa_{11}+\frac{9(33n-82)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{21}\kappa_{11}$$
$$+\frac{54}{(n-1)^2}\kappa_{13}\kappa_{30}\kappa_{11}+\frac{18}{(n-1)^2}\kappa_{31}\kappa_{03}\kappa_{20}+\frac{36(3n-7)}{(n-1)^2(n-2)}\kappa_{22}\kappa_{12}\kappa_{20}$$
$$+\frac{18(3n-8)}{(n-1)^2(n-2)}\kappa_{13}\kappa_{21}\kappa_{20}+\frac{9}{(n-1)^2}\kappa_{40}\kappa_{12}\kappa_{02}+\frac{9}{(n-1)^2}\kappa_{40}\kappa_{03}\kappa_{11}$$
$$+\frac{27}{(n-1)^2}\kappa_{30}\kappa_{03}\kappa_{21}+\frac{9(19n^2-84n+100)}{(n-1)^2(n-2)^2}\kappa_{21}^2\kappa_{12}+\frac{54}{(n-1)^2}\kappa_{12}^2\kappa_{30}$$
$$+\frac{18n}{(n-1)^2(n-2)}\kappa_{03}\kappa_{20}^2\kappa_{11}+\frac{36n(3n-8)}{(n-1)^2(n-2)^2}\kappa_{20}\kappa_{21}\kappa_{11}\kappa_{02}$$
$$+\frac{18n}{(n-1)^2(n-2)}\kappa_{12}\kappa_{20}^2\kappa_{02}+\frac{18n(9n-22)}{(n-1)^2(n-2)^2}\kappa_{21}\kappa_{11}^3+\frac{36n(5n-12)}{(n-1)^2(n-2)^2}\kappa_{12}\kappa_{20}\kappa_{11}^2$$
$$+\frac{54n}{(n-1)^2(n-2)}\kappa_{30}\kappa_{11}^2\kappa_{02}.$$

$$\kappa\begin{pmatrix}3&2\\0&0\end{pmatrix}=\frac{1}{n}\kappa_{50}+\frac{6}{n-1}\kappa_{30}\kappa_{20},$$
$$\kappa\begin{pmatrix}3&1\\0&1\end{pmatrix}=\frac{1}{n}\kappa_{41}+\frac{3}{n-1}\kappa_{30}\kappa_{11}+\frac{3}{n-1}\kappa_{21}\kappa_{20},$$
$$\kappa\begin{pmatrix}2&2\\1&0\end{pmatrix}=\frac{1}{n}\kappa_{41}+\frac{2}{n-1}\kappa_{30}\kappa_{11}+\frac{4}{n-1}\kappa_{21}\kappa_{20},$$
$$\kappa\begin{pmatrix}3&0\\0&2\end{pmatrix}=\frac{1}{n}\kappa_{32}+\frac{6}{n-1}\kappa_{21}\kappa_{11},$$
$$\kappa\begin{pmatrix}2&1\\1&1\end{pmatrix}=\frac{1}{n}\kappa_{32}+\frac{3}{n-1}\kappa_{21}\kappa_{11}+\frac{2}{n-1}\kappa_{12}\kappa_{20}+\frac{1}{n-1}\kappa_{30}\kappa_{02},$$
$$\kappa\begin{pmatrix}1&2\\2&0\end{pmatrix}=\frac{1}{n}\kappa_{32}+\frac{4}{n-1}\kappa_{21}\kappa_{11}+\frac{2}{n-1}\kappa_{12}\kappa_{20}.$$

I am indebted to Dr F. N. David and Dr John Wishart for a number of suggestions and help in the presentation of this paper.

REFERENCES

David, F. N. & Kendall, M. G. (1949). *Biometrika*, **36**, 431.
Fisher, R. A. (1929). *Proc. Lond. Math. Soc.* **30**, 199.
Kendall, M. G. (1940). *Ann. Eugen., Lond.*, **10**, 392.
Wishart, J. (1928). *Proc. Lond. Math. Soc.* **29**, 309.

RANDOM DISPERSAL IN THEORETICAL POPULATIONS

By J. G. SKELLAM

The Nature Conservancy, London

Synopsis

The random-walk problem is adopted as a starting point for the analytical study of dispersal in living organisms. The solution is used as a basis for the study of the expansion of a growing population, and illustrative examples are given. The law of diffusion is deduced and applied to the understanding of the spatial distribution of population density in both linear and two-dimensional habitats on various assumptions as to the mode of population growth or decline. For the numerical solution of certain cases an iterative process is described and a short table of a new function is given. The equilibrium states of the various analytical models are considered in relation to the size of the habitat, and questions of stability are investigated. A mode of population growth resulting from the random scattering of the reproductive units in a population discrete in time, is deduced and used as a basis for a study on interspecific competition. The extent to which the present analytical formulation is applicable to biological situations, and some of the more important biological implications are briefly considered.

1. Introduction

1·1. It is now fifty years since the publication of *The Origin of the British Flora* by Clement Reid (1899). In it is suggested an interesting numerical problem on the rate of dispersal of plants. Reid states: 'Though the post-glacial period counts its thousands of years, it was not indefinitely long, and few plants that merely scatter their seed could advance more than a yard in a year, for though the seed might be thrown further, it would be several seasons before an oak for instance, would be sufficiently grown to form a fresh starting point. The oak, to gain its present most northerly position in North Britain after being driven out by the cold, probably had to travel fully six hundred miles, and this without external aid would take something like a million years.'

1·2. At the end of the last century, biologists, unlike physicists, rarely formulated such problems in terms of simplified abstract models, due no doubt to the comparatively greater complexity of biological systems. A beginning might have been made on the subject of dispersal, for much of the necessary mathematical technique had been developed already, and, in fact, had been utilized by Maxwell (1860) in developing a kinetic theory of gases based on the behaviour of an infinity of perfectly elastic spheres moving at random. The present century has witnessed the great success of the analytical method to quote only the work of Fisher (1930), Haldane (1932) and Wright (1931) in evolutionary genetics, and of Volterra (1931), Lotka (1925, 1939) and Kostitzin (1939) in ecology. Nevertheless, biologists as a whole have been reluctant to develop the analytical as distinct from the purely statistical approach, and apart from the pioneer work of Karl Pearson (1906) and of Brownlee (1911), the mathematical aspects of the problem of dispersal have not received the attention they deserve.

2. RANDOM DISPERSAL AND DIFFUSION

2·1. *Random displacement.* The mathematical methods employed in the present treatment are largely the natural outcome of the solution of the well-known random-walk problem associated with Pearson (1906), Rayleigh (1919) and Kluyver (1905), and reviewed more recently by Chandrasekhar (1943). In the simplest form of this problem, a particle on a line jumps one place to the right or one to the left, both being equally likely. It is easily seen that the probability distribution is:

Position	$\bar{3}$	$\bar{2}$	$\bar{1}$	0	1	2	3
After 1 jump	.	.	$\frac{1}{2}$	.	$\frac{1}{2}$	.	.
After 2 jumps	.	$\frac{1}{4}$	.	$\frac{2}{4}$	.	$\frac{1}{4}$	.
After 3 jumps	$\frac{1}{8}$	.	$\frac{3}{8}$	.	$\frac{3}{8}$	.	$\frac{1}{8}$

and so on, with the general result that after n jumps the distribution is binomial with variance equal to n. As n becomes large the distribution 'tends to' normality.

A natural extension of the problem to N dimensions, in which the particle moves on a rectangular lattice, yields 'in the limit' an N-variate normal distribution which is the product of N independent distributions.

An alternative model for the two- or three-dimensional analogue consists of a randomly kinked chain with rotatable links. If one end is fixed, the probability distribution of the position of the free end tends to normality when the number of links is large (see Uspensky, 1937). In the two-dimensional case with only seven links it has been shown by actual calculation by Pearson (1906) that the normal distribution provides a remarkably good approximation.

The model considered by Pearson is, however, somewhat artificial when applied to migration in that the links (or flights) are taken as being equal in size. In practice they will vary. The important special case where the probability distribution of the end-point of a flight is bivariate-normal (with the starting point of the flight as the centre) has been considered by Brownlee (1911). A solution of the more general case is outlined below.

2·2. *A spreading population.* Consider a surface of unlimited extent approximating to a Euclidean plane with co-ordinate axes OX and OY, and in the immediate neighbourhood of the origin let there be to start with a single self-reproducing particle of a species not otherwise represented. The probability distribution of the position of a particular descendant after n generations may be deduced once the characteristic function of the probability distribution of offspring about their parent is given. For a parent located at (ξ, η), let this function be

$$\phi(t, u \mid \xi, \eta) = \exp\{i\xi t + i\eta u + \Sigma(\kappa_{rs} t^r u^s i^{r+s}/r\,!\,s\,!)\}.*$$

The κ_{rs} are assumed independent of ξ and η, for it is reasonable to suppose that the location of the parent does not affect the *form* of the distribution of the offspring. Let the cumulative distribution function for the position of a particle in the nth generation be $F_n(\xi, \eta)$ with $\phi_n(t, u)$ as the corresponding characteristic function. Assuming that the generations do not overlap in time, we then have

$$\phi_{n+1}(t, u) = \iint \phi(t, u \mid \xi, \eta)\, dF_n(\xi, \eta) = \phi(t, u \mid 0, 0) \iint e^{i\xi t + i\eta u}\, dF_n(\xi, \eta) = \phi(t, u \mid 0, 0)\, \phi_n(t, u).$$

* Here and below in this paper the customary parentheses surrounding terms in the denominator, to the right of the solidus, have been omitted. Thus ... $/r\,!\,s\,!\}$ should be printed ... $/(r\,!\,s\,!)\}$ in accordance with the common convention. It is considered, however, that no confusion will result here from such omissions.

Clearly $\phi_n = [\phi(t, u \mid 0, 0)]^n$. It is apparent that all cumulants increase n-fold (as in the summation of independent variables). Standardization of the scale is effected by changing $n\kappa_{rs}$ to $n\kappa_{rs}/n^{\frac{1}{2}(r+s)}$. The approach to normality is obviously very rapid when the distribution of offspring about their parent is roughly normal to begin with. In the present treatment we assume that $\kappa_{01} = \kappa_{10} = 0$ so that there is no drift. Nevertheless, it should be noted that a slight systematic drift, no matter how small, is *ultimately* the most important cause of displacement, that is, when n is chosen sufficiently large.

The distribution of the position of a particle of the nth generation will henceforth be taken as

$$dF_n(x, y) = \pi^{-1}a^{-2}n^{-1}\exp\{-[x^2+y^2]/na^2\}\,dx\,dy. \tag{1}$$

Since we shall be concerned mainly with absolute distances from the origin, it is convenient to make the polar transformation $x = r\cos\theta$, $y = r\sin\theta$. From (1) we then obtain

$$dF_n = \pi^{-1}a^{-2}n^{-1}\exp\{-r^2/na^2\}\,r\,dr\,d\theta \quad (0 \leqslant \theta < 2\pi, 0 \leqslant r < \infty). \tag{2}$$

If r_0 is the distance from the origin of an individual value chosen at random, then the probability that $r - \frac{1}{2}dr \leqslant r_0 < r + \frac{1}{2}dr$ is obtained by integrating over θ, giving the radial probability density

$$f(r \mid n, a^2) = \exp\{-r^2/na^2\}\,2r/na^2 \quad (0 \leqslant r < \infty). \tag{3}$$

The parameter a^2 is to be interpreted as the mean-square dispersion per generation, analogous with the mean-square velocity in Maxwell's distribution. The maximum likelihood estimate of a^2 based on ν observed values of r is simply $\hat{a^2} = \Sigma r^2/n\nu$.

Of the population spread out after n generations, that proportion lying outside a circle of radius R is

$$p = \int_R^\infty \exp[-r^2/na^2]\,2r\,dr/na^2 = \exp\{-R^2/na^2\}. \tag{4}$$

In a subsequent argument it is essential to make p very small indeed. By choosing $p = 1/N_n$, where N_n is the final population after n generations, it follows that only one particle can be expected outside a circle of radius $R_n = (na^2 \log N_n)^{\frac{1}{2}}$. If during this period the population has increased geometrically in accordance with the relation $N = e^{n\gamma^2}$ (where $c = \gamma^2$ may be termed the Malthusian parameter, as in Fisher (1930)), we obtain immediately the relation

$$R_n/n = a\gamma. \tag{5}$$

If a circular contour is drawn in the XY plane through all points having an arbitrarily chosen constant areal population density, then as time passes, the contour expands outwards. From the condition, $\gamma^2 n - \log n - r^2/na^2 = \text{constant}$, it is easily seen that

$$r_{n+1} - r_n \sim a\gamma. \tag{5a}$$

The ultimate velocity of propagation of this 'wave-front' (D. G. Kendall, 1948) is therefore the same as the constant velocity of expansion of the circle within which all but one of the population can be 'expected' to lie.

2·3. *A numerical illustration.* Let us consider the application of equation (4) to Reid's problem. We can clearly establish a rigorous conclusion in the form of an inequality provided that we can fix appropriate bounds to the various parameters.

The oak does not produce acorns until it is sixty or seventy years old, and even then it is not mature. It then produces acorns over a period of several hundred years. The rate of population expansion is obviously very much less than it would be if all the immediate offspring of a single individual could come into existence in the sixtieth year of its life. We

are certainly safe in taking the generation time as not less than sixty years. Now according to De Geer (1910) the final recession of the ice started at about 18,000 B.C., and yet in Roman Britain (Tansley, 1939, p. 171) oak forests were apparently well established throughout the country. It appears safe to take $n \leqslant 300$.

The number of acorns produced by an oak in the course of its life must be prodigious. Nevertheless, under typical present-day conditions almost every acorn which falls to the ground in autumn is consumed by birds and mammals (Watt, 1919). Of those remaining, some decay and many fail to germinate. Mortality among the young seedlings is very high, and even when not overshadowed it seems that only about 1 % of the seedlings are likely to survive the next three years. Though there has been much speculation, little is really known about the biological conditions following the last glaciation. Even so, it seems perfectly safe to assume that the average number of mature daughter oaks produced by a single parent oak did not exceed 9 million. To most biologists such a figure must appear unnecessarily high, but we must remember that it is the rate of population growth at the periphery of the advancing 'wave' which matters most, and here there will be least competition.

The fact that the British Isles are of finite extent with irregular coastline does not detract from the power of the argument, for it is obvious that such a shape in no way encourages the advancing 'wave-front'.

We then have $R/a < 300\sqrt{(\log 9{,}000{,}000)} = 1200$.

In the original form of the problem as stated by Reid, R is given as 600 miles. It then follows that a (the root mean square distance of daughter oaks about their parent) $> \frac{1}{2}$ mile. On these premises the conclusion which Reid reached appears inescapable—namely, that animals such as rooks must have played a major role as agents of dispersal.

If, however, we regard the upper bounds of a, n, γ (given above) as collectively excessive, we are driven to accept the view that the last glaciation was not as extensive as was previously believed (W. B. Wright, 1937), and to suppose that the oak population regenerated from scattered pockets which survived in favourable valleys.

At this point it is interesting to consider the actual succession of forest trees since glacial times as revealed by pollen analysis (Godwin, 1934). In late sub-arctic and pre-boreal times, birches with their small, winged fruits, willows with their plumed seeds, and pine with its winged seeds, were able to spread quickly and were plentiful. Hazel, however, did not reach its maximum until Boreal times, and oak with still larger fruits not until much later. It is quite possible that the relatively different rates of dispersal were not unimportant in determining the order of succession.

2·4. *The dispersal of small animals.* The results already deduced can be applied equally well to the dispersal of small animals such as earthworms and snails. For example, if the mean square 'dispersion' per minute of a wingless ground beetle wandering at random is 1 yard^2, then after a season of 6 months, say, without rest, the R.M.S.D. of the resulting probability distribution is only 500 yards. The probability that after 6 months the beetle wanders more than a mile from the starting point is less than 8 in a million. Without external aid a period of time equivalent to 1,000,000 seasons would be required to raise the R.M.S.D. to 290 miles. The figures, however, are slightly deceptive, for the final position arrived at is in general a great deal nearer the origin than the farthermost position previously reached. Nevertheless, calculations of this kind clearly indicate in these cases the importance in the long run of rare accidental displacements due to external agencies such as high gales, floating plant material, the muddy feet of birds and mammals.

2·5. *Empirical confirmation.* In practice there is rarely sufficient information to construct the contours of population density with accuracy. One contour, however, can sometimes be drawn—that for the low 'threshold' density (depending on the thoroughness of the survey) at which the population begins to escape notice altogether.

Equation (4), derived initially on theoretical grounds, is well illustrated by the spread of the muskrat, *Ondatra zibethica* L., in central Europe since its introduction in 1905. Fig. 1, based on Ulbrich (1930), shows the apparent boundaries for certain years. If we are prepared to accept such a boundary as being representative of a theoretical contour, then we must regard the area enclosed by that boundary as an estimate of πr^2. The relation between the time and $\sqrt{\text{area}}$ is shown graphically in Fig. 2.

Fig. 1

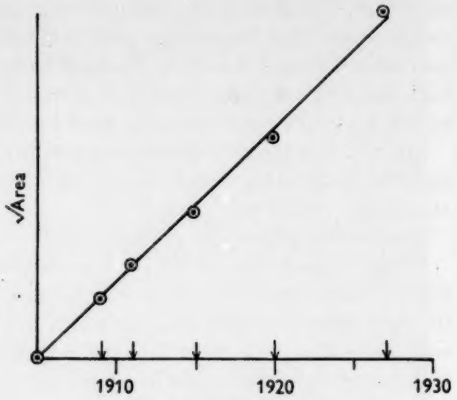


Fig. 2

2·6. *The analogy with diffusion.* If a random particle suffers a displacement ϵ in any direction at regular intervals of time $(t, t+\omega, t+2\omega, \ldots)$, and if the probability density is denoted by ψ, it is clear that $\psi(x, y, t+\omega)$ is the mean value of $\psi(\xi, \eta, t)$ for all (ξ, η) on a circle of radius ϵ around (x, y). That is,

$$\psi(x, y, t+\omega) = \frac{1}{2\pi}\int_0^{2\pi} \psi(x+\epsilon\cos\theta, y+\epsilon\sin\theta, t)\, d\theta.$$

Expanding by Taylor's theorem and noting that

$$\int_0^{2\pi} \cos\theta\, d\theta = \int_0^{2\pi} \sin\theta\, d\theta = \int_0^{2\pi} \sin\theta\cos\theta\, d\theta = 0,$$

$$\int_0^{2\pi} \cos^2\theta\, d\theta = \int_0^{2\pi} \sin^2\theta\, d\theta = \pi,$$

we obtain for infinitesimally small ϵ and ω the relation

$$\frac{\partial\psi}{\partial t} = \frac{1}{4}\frac{\epsilon^2}{\omega}\nabla^2\psi, \tag{6}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

The partial differential equation (6) is of course well known in connexion with the conduction of heat and related problems in mathematical physics (Jeans, 1921). The one-dimensional form has been applied by R. A. Fisher (1922) to the problem of the distribution of gene differences, and the three-dimensional form by Rashevsky (1948) to problems arising from the diffusion of materials from and into living cells.

It should be noted that the fundamental differential equation (6) is satisfied by the distribution

$$\psi = \pi^{-1}a^{-2}t^{-1}\exp\{-(x^2+y^2)/ta^2\},$$

already considered (see equation (1)), with the unit of time as one generation and $a^2 = \epsilon^2/\omega$.

It is apparent that many ecological problems have a physical analogue, and that the solution of these problems will require treatments and the use of functions with which we are already very familiar. Unlike most of the particles considered by physicists, however, living organisms reproduce, and members of the same and of different species interact. As a result the equations of mathematical ecology are often of a new and unusual kind (see, for example, Volterra (1931)), and require special treatment.

3. DISPERSAL AND POPULATION DENSITY

3·1. *Modes of population growth.* Many problems on dispersal cannot be formulated unless some law of population growth (in the absence of dispersal) is assumed. As long as the population is small or shows a natural tendency to decrease, the Malthusian law $dN/dt = cN$ is usually satisfactory. If the population is not small the Pearl-Verhulst logistic law is more appropriate.

This law may be written in the form

$$\frac{dN}{dt} = cN - lN^2, \tag{7}$$

where, following Kostitzin (1939), c may be termed the coefficient of increase and l the limiting coefficient. Since a stable population level exists at $M \equiv c/l$, the law is often written in the form

$$\frac{dN}{dt} = cN(M-N)/M. \tag{7a}$$

For our present purpose it is convenient to use the concept of population density denoted by Ψ, a function of time and place. In general, the vital coefficients will also vary. The logistic law may then be written

$$\frac{\partial}{\partial t}\Psi(x,y,t) = c_1(x,y)\Psi - c_2(x,y)\Psi^2. \tag{8}$$

In the simplest cases where the vital coefficients remain constant throughout the habitat under consideration, the unit of population density may be chosen for convenience as that prevailing when $t \to \infty$, and the law then takes the very simple form

$$\frac{d\Psi}{dt} = c\Psi(1-\Psi). \tag{9}$$

When conditions become extremely unfavourable and ultimate survival not possible, the logistic law (7) may still be applied with c negative ($= -g^2$). The population declines rapidly unless maintained by immigration. Such a population may be termed negative-logistic.

A convenient change of scale then gives

$$\frac{d\Psi}{dt} = -g^2\Psi(1+\Psi).$$

Two further laws of population growth applicable to populations which are discrete in time are deduced in §§4·1 and 4·3.

3·2. *Centres of multiplication and extinction.* We now consider the distribution in space of the density of a population which by reason of random dispersal, tends to flow from regions in which conditions are favourable to regions in which they are unfavourable. Whether a steady state can exist or not depends on the ability of the population in the favourable regions to make good the decline in the unfavourable ones.

Mathematically it is simpler to consider a linear habitat such as a coastal zone (Fisher, 1937) or the bank of a river. In these cases we are justified in assuming that an adjustment has been made in the vital coefficients to allow for diffusion inland, for this danger may be regarded as one of the many hazards associated with the habitat. Moreover, as will be apparent later, the mathematical form of the Malthusian law and of the logistic law (7) persists almost unchanged when the rate of diffusion inland is deducted from the rate of population growth.

The two-dimensional problems which most readily lend themselves to mathematical treatment are those with radial symmetry. Rough approximations to circles are sometimes afforded by islands, hill tops, marshes, patches of woodland, and approximations to annuli are provided at the margins of lakes, by the zones on hills of a conical or hemispherical shape, or even by the zonation created by rabbits around their individual burrows or around such patches of woodland as they infest (see, for example, Tansley, 1939, pp. 587, 505, 138, 141).

3·3. *Malthusian populations in linear habitats.* The one-dimensional form of (6) expressing the effect of random dispersal is

$$\frac{\partial\psi}{\partial t} = \frac{1}{2}\frac{\epsilon^2}{\omega}\frac{\partial^2\psi}{\partial x^2}. \tag{10}$$

In terms of population density, with $a^2 = \epsilon^2/\omega$, we then have as the combined effect of dispersal and population growth

$$\frac{\partial\psi}{\partial t} = \tfrac{1}{2}a^2\frac{\partial^2\Psi}{\partial x^2} + c\Psi. \tag{11}$$

The condition for a steady state, assuming it to exist, is

$$\frac{d^2\Psi}{dx^2} + \frac{2c}{a^2}\Psi = 0. \tag{12}$$

Case (*i*). Bordering on a suitable habitat is an unfavourable zone extending indefinitely in one direction:

$$c = -g^2 \quad \text{in} \quad x_b \leqslant x < \infty, \quad \Psi(x_b) = B, \quad \Psi(\infty) = 0.$$

Solution: $$\Psi = Be^{-m(x-x_b)}, \quad \text{where} \quad m = g\sqrt{2}/a.$$

Case (*ii*). An unfavourable interval is flanked at both ends by a suitable habitat:

$$c = -g^2 \quad \text{in} \quad -x_b \leqslant x \leqslant x_b, \quad \Psi(-x_b) = \Psi(x_b) = B.$$

Solution: $$\Psi = B\cosh mx/\cosh mx_b.$$

Case (*iii*). An apparently suitable habitat is flanked on either side by conditions which are *extremely* unfavourable. (It is assumed, for simplicity, that the organisms show no marked negative tactic response to the changed conditions, so that there is no appreciable reflexion at the boundary):

$$c = \gamma^2 \quad \text{in} \quad -x_b \leqslant x \leqslant x_b, \quad \Psi(-x_b) = \Psi(x_b) = 0.$$

Solutions that are non-negative but not zero everywhere in the interval are possible only when $x_b = \frac{1}{2}\pi a/\gamma\sqrt{2}$, and we then have

$$\Psi = C\cos(x\gamma\sqrt{2}/a). \tag{13}$$

When x_b is greater than the critical value given above, the population increases indefinitely, and when x_b is less than this value the population declines to extinction. For consider the partial differential equation (11) with $c = \gamma^2$ subject to the boundary conditions

$$\Psi(x_b, t) = \Psi(-x_b, t) = 0 \quad (0 \leqslant t < \infty),$$

and an arbitrary initial state developed as a Fourier half-range sine series

$$\Psi(x, 0) = \sum_{s=1}^{\infty} A_s \sin(\tfrac{1}{2}s\pi[x + x_b]/x_b).$$

The solution is

$$\Psi(x, t) = \sum_s A_s c^{k_s t} \sin(\tfrac{1}{2}s\pi[x + x_b]/x_b),$$

where $k_s = \gamma^2 - \frac{1}{2}a^2(\frac{1}{2}s\pi/x_b)^2$. The first of the k_s is obviously the greatest, so that in the course of time the solution is dominated by its first term, which tends to ∞, (13), or 0 as $x_b \gtrless \frac{1}{2}\pi a/\gamma\sqrt{2}$.

3·4. *Logistic populations in linear habitats.* From (10) and (9) we obtain the fundamental equation

$$\frac{\partial\Psi}{\partial t} = \tfrac{1}{2}a^2\frac{\partial^2\Psi}{\partial x^2} + \gamma^2\Psi(1 - \Psi). \tag{14}$$

Mathematically this is the same as an equation of R. A. Fisher (1937) given as

$$\frac{\partial p}{\partial t} = k\frac{\partial^2 p}{\partial x^2} + mpq$$

in connexion with the problem of gene flow in a linear population. Gene frequency p corresponds to population density and the term in $pq \equiv p(1-p)$ expressing the effect of natural selection corresponds to the term in $\Psi(1-\Psi)$ expressing logistic growth. The results deduced here in connexion with ecological problems automatically apply to the genetic analogues where they exist. Whereas, however, Ψ may exceed unity, p is restricted to $0 \leqslant p \leqslant 1$. Negative solutions are inadmissible in both types of problem.

The condition for a steady state is

$$\frac{d^2\Psi}{dz^2} + \Psi(1 - \Psi) = 0, \quad \text{where} \quad z = x\gamma\sqrt{2}/a, \tag{15}$$

or in the case of a negative-logistic population (see (3·1))

$$\frac{d^2\Psi}{dz^2} - \Psi(1 + \Psi) = 0, \quad \text{where} \quad z = xg\sqrt{2}/a. \tag{16}$$

The substitution $u = (\Psi_0 - \Psi)/\Psi_0(1 - \Psi_0)$, where $\Psi_0 \equiv \Psi(x)|_{x=0}$,

converts (15) into $u'' = 1 + \lambda u - \frac{1}{4}(1 - \lambda^2)u^2$, where $\lambda = 2\Psi_0 - 1$. (17)

Case (*i*). $$c = \gamma^2, \quad 0 < \Psi_0 < 1, \quad \left.\frac{d\Psi}{dx}\right|_{x=0} = 0.$$

Since $\frac{d^2u}{dz^2} = \frac{du}{dz}\frac{d}{du}\left(\frac{du}{dz}\right)$ we find on integrating (17), subject to the condition $du/dz = 0$ when $u = 0$, that

$$\frac{1}{2}\left(\frac{du}{dz}\right)^2 = u + \tfrac{1}{2}\lambda u^2 - \tfrac{1}{12}(1-\lambda^2)\,u^3.$$

Hence
$$|z|\sqrt{2} = \int_0^u \{\tau(1-\alpha^2\tau)(1+\beta^2\tau)\}^{-\frac{1}{2}}\,d\tau,$$

where
$$4\alpha^2 = \left(\frac{4-\lambda^2}{3}\right)^{\frac{1}{2}} - \lambda \quad \text{and} \quad 4\beta^2 = \left(\frac{4-\lambda^2}{3}\right)^{\frac{1}{2}} + \lambda.$$

In terms of Legendre's elliptic integrals, of which there are tables (1825), we then have the solution

$$z\rho 2^{-\frac{1}{2}} = F(k, \tfrac{1}{2}\pi) - F(k, \arccos\alpha\sqrt{u}),$$

where
$$\rho^2 = \alpha^2 + \beta^2 \quad \text{and} \quad k^2 = 2\beta^2 \Big/ \left(\frac{4-\lambda^2}{3}\right)^{\frac{1}{2}}.$$

The inversion of this result leads to a simpler expression in terms of the Jacobian elliptic function sd (y) denoting $\sin \operatorname{am} y/\Delta \operatorname{am} y$. We thereby write

$$\sqrt{u} = \operatorname{sd}(2^{-\frac{1}{2}}\rho z, k)/\rho. \tag{18}$$

In practice, however, it is no more troublesome to compute a particular solution by iteration. For under the condition $\left.\frac{d\Psi}{dz}\right|_{z=0} = 0$ we may write equation (15) in the form

$$\Psi(z) = \Psi(0) - \int_0^z d\tau \int_0^\tau \Psi(\zeta)\,[1-\Psi(\zeta)]\,d\zeta.$$

Approximate values of Ψ are substituted on the right and the integrations performed numerically as in §3·9.

If we make Ψ_0 vanishingly small and proceed to determine z_b such that $\Psi(z_b) = 0$, we find that $u(z_b) \to 1$ and $\rho \to 2^{-\frac{1}{2}}$. Equation (18) passes formally into $2^{-\frac{1}{2}} = \sin \frac{1}{2}z_b$, confirming the result $x_b = \frac{1}{2}\pi a/\gamma\sqrt{2}$ given in §3·3 (iii).

Case (*ii*). $$c = \gamma^2, \quad 1 < \Psi_0 < \tfrac{3}{2}, \quad \left.\frac{d\Psi}{dx}\right|_{x=0} = 0.$$

Following the same procedure as before we find

$$|z|\sqrt{2} = \int_0^u \{\tau(1+\alpha^2\tau)(1+\beta^2\tau)\}^{-\frac{1}{2}}\,d\tau,$$

where
$$4\alpha^2 = \lambda - \left(\frac{4-\lambda^2}{3}\right)^{\frac{1}{2}} \quad \text{and} \quad 4\beta^2 = \lambda + \left(\frac{4-\lambda^2}{3}\right)^{\frac{1}{2}}.$$

Hence
$$2^{-\frac{1}{2}}\beta z = F(k, \arctan\beta\sqrt{u}), \quad \text{where} \quad k = \left(\frac{4-\lambda^2}{3}\right)^{\frac{1}{2}} \Big/ \beta\sqrt{2}. \tag{19}$$

Alternatively
$$\sqrt{u} = \operatorname{tn}(2^{-\frac{1}{2}}\beta z, k)/\beta.$$

Since $$F(k,\theta) \equiv \int_0^{\sin\theta} \{(1-t^2)(1-k^2t)\}^{-\frac{1}{2}}\,dt,$$

$$F(1,\theta) = \arg\tanh(\sin\theta) = gd^{-1}\theta.$$

In the important case when Ψ_0 is close to 1, so that β^2 is near $\frac{1}{2}$ and k near 1, we find that (19) passes into

$$\tfrac{1}{2}z = gd^{-1}(\arctan\sqrt{(\tfrac{1}{2}u)}) \quad \text{or} \quad \sinh\tfrac{1}{2}z = \sqrt{(\tfrac{1}{2}u)},$$

with the result that $$\cosh z = 1+u = (\Psi-1)/(\Psi_0-1),$$

a conclusion which may be conjectured directly from the differential equation itself.

Other cases. Solutions in terms of elliptic functions are obtained in a similar manner after making appropriate substitutions. No special treatment is required in the case of a negative logistic population maintained by immigration in a region of extinction. We merely write $\Psi+1$ instead of Ψ in the corresponding solution for the positive logistic case.

3·5. *Malthusian populations in two-dimensional habitats.* The fundamental equation expressing the combined effect of diffusion and geometric increase in a habitat throughout which conditions are constant is

$$\frac{\partial\Psi}{\partial t} = \tfrac{1}{4}a^2\nabla^2\Psi + c\Psi(x,y,t). \tag{20}$$

In radially symmetrical cases we may write

$$\frac{\partial\Psi}{\partial t} = \frac{1}{4}a^2\left(\frac{\partial^2\Psi}{\partial r^2} + \frac{1}{r}\frac{\partial\Psi}{\partial r}\right) + c\Psi(r,t), \tag{21}$$

where $r^2 = x^2+y^2$.

Case (*i*). A circular region of multiplication enclosed in a zone of *absolute* extinction:

$$c = \gamma^2 \quad \text{in} \quad 0 \leqslant r < r_b, \quad 0 \leqslant t < \infty; \quad \Psi(r_b,t) = 0.$$

To ensure that the solution is finite at the origin we impose the further condition $\dfrac{\partial}{\partial r}\Psi\Big|_{r=0} = 0$.

The arbitrary initial state may be expressed as a Bessel series

$$\Psi(r,0) = \sum_s A_s J_0(rj_s/r_b),$$

where j_s is the sth zero of J_0 and A_s is determined in the usual Fourier manner by reason of the orthogonality relation

$$\int_0^1 yJ_0(yj_s)\,J_0(yj_\sigma)\,dy = 0 \quad (s \neq \sigma).$$

The solution is $$\Psi(r,t) = \sum_s A_s e^{k_s t} J_0(rj_s/r_b),$$

where $$k_s = \gamma^2 - \tfrac{1}{4}a^2j_s^2/r_b.$$

Clearly $\Psi \sim A_1 e^{k_1 t} J_0(rj_1/r_b)$, tending to 0 or ∞ as $r_b \lesseqgtr j_1 a/2\gamma$. It appears again that habitats below a certain critical size, depending on the conditions, are insufficient to maintain a species undergoing unrestricted random dispersal.

Case (*ii*). Let (ξ,η) be the co-ordinates of the place of origin of an individual of the parental generation for which the density function in space is $\Psi(\xi,\eta)$. Let the probability that a particular one of its offspring originates in the interval $(x \pm \frac{1}{2}dx;\ y \pm \frac{1}{2}dy)$ be

$$\alpha(x-\xi, y-\eta \mid a^2)\,dx\,dy.$$

Let the probability that such an offspring reaches maturity be $P(x, y)$. For most simple, small populations P may be taken independent of Ψ. Then if on the average the number of original offspring per parent is ν we may define a viability function $V(x, y) \equiv \nu P(x, y)$. The condition for a steady state from generation to generation is the integral equation

$$\Psi(x, y) = V(x, y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(\xi, \eta)\, \alpha(x-\xi, y-\eta \mid a^2)\, d\xi\, d\eta. \tag{22}$$

For the purpose of the present simplified model we take α to be the function

$$\pi^{-1} a^{-2} \exp\{-[(x-\xi)^2 + (y-\eta)^2]/a^2\},$$

a choice justified by previous arguments. The function $V(x, y)$ needs to be positive for all values of the variable, and preferably continuous. It will then be seen that the choice $V(x, y) = \exp\{(b^2 - x^2 - y^2)/v\}$ not only gives a reasonable spatial distribution of viability, but also renders the integral equation tractable. Viability is greatest at the origin, $V_0 = \exp\{b^2/v\}$. The circle $x^2 + y^2 = b^2$ divides the plane into two parts, and in the absence of dispersal, the population inside the circle would increase and the population outside would decline.

The integral equation (22) then has a solution of the form

$$\Psi(\xi, \eta) = A \pi^{-1} \theta^{-1} \exp\{-(\xi^2 + \eta^2)/\theta\}. \tag{23}$$

Since the double integral is the product of independent integrals and since

$$\int_{-\infty}^{\infty} \alpha(\xi \mid \theta_1)\, \alpha(x-\xi \mid \theta_2)\, d\xi = \alpha(x \mid \theta_1 + \theta_2),$$

where α is a normal function, we find on substituting (23) into (22) and equating coefficients that

$$\frac{1}{\theta} = \frac{1}{v} + \frac{1}{\theta + a^2} \quad \text{and} \quad \frac{1}{\theta} = \frac{V_0}{\theta + a^2}, \quad \text{where} \quad \log V_0 = b^2/v.$$

From these it follows that $\qquad b^2 = a^2 V_0 \log V_0 / (V_0 - 1)^2$.

In the sense indicated earlier, b is the radius of the region of multiplication. Unless this value is attained the population must decline to extinction.

Case (*iii*). An infinite zone of extinction around a single circular area of multiplication:

$$c = -g^2 \quad \text{in} \quad r_b \leqslant r < \infty, \quad \Psi(r_b) = B, \quad \Psi(\infty) = 0.$$

The steady state is given by

$$\frac{d^2\Psi}{dr^2} + \frac{1}{r}\frac{d\Psi}{dr} - m^2 \Psi = 0 \quad \text{where} \quad m = 2g/a. \tag{24}$$

The solution is $\qquad \Psi(r) = A \int_0^{\infty} e^{-rm \cosh u}\, du = A K_0(rm),$

where K_0 is the modified Bessel function of the second kind of order zero, and $A = B/K_0(r_b m)$. It may be noted that $K_0(z) \sim e^{-z}(\pi/2z)^{\frac{1}{2}}$ for large z. When r_b is very large the problem reduces to a linear one (§3·3 (i)), the radial component of the R.M.S.D./generation being $a/\sqrt{2}$.

As a numerical illustration suppose that in the outer zone the population shows a natural tendency to decline at a rate of very nearly 4 % per generation ($g = 0{\cdot}2$). If the radius of the reservoir is 1000 yards and the R.M.S.D./generation 100 yards, then the population density

at a point 1000 yards outside the boundary compared with that at the boundary will be approximately $e^{-2000m}\,2000^{-\frac{1}{2}}/(e^{-1000m}\,1000^{-\frac{1}{2}}) = 0{\cdot}01295$, in close agreement with the more accurate value 0·01312 based on the tables of K_0 given in Watson (1944).

Case (*iv*). A circular region of extinction:

$$c = -g^2 \quad \text{in} \quad 0 \leqslant r \leqslant r_b, \quad \Psi'(0) = 0, \quad \Psi(r_b) = B.$$

Solution: $$\Psi(r) = BI_0(rm)/I_0(r_b m),$$

where I_0 is the modified Bessel function of the first kind of order zero.

Case (*v*). An annulus of multiplication in a region of *absolute* extinction:

$$c = \gamma^2 \quad \text{in} \quad r_a \leqslant r \leqslant r_b, \quad \Psi(r_a) = \Psi(r_b) = 0.$$

A stationary non-negative solution exists for every fixed value r_a provided that the value of r_b is made to depend on r_a. If μ denotes $2\gamma/a$ the required solution is

$$\Psi(r) = A[J_0(\mu r)\,Y_0(\mu r_a) - J_0(\mu r_a)\,Y_0(\mu r)],$$

where Y_0 is the Bessel function of the second kind (of Weber's type) of order zero.

If $h = r_b/r_a$, it may be seen that μr_a is the first root (x_0) of the equation

$$J_0(hx)\,Y_0(x) - J_0(x)\,Y_0(hx) = 0. \tag{25}$$

The relation between r_b/r_a and $\mu r_b - \mu r_a$ is that between h and $(h-1)x_0$, the form in which the zeros of (25) are tabulated in Jahnke & Emde (1945). As r_a increases from zero, $\mu(r_b - r_a)$ moves from j_{01} to π, and the problem rapidly degenerates to the linear case.

3·6. *Logistic populations in two-dimensional habitats.* The combined effect of diffusion and logistic growth in radially symmetrical habitats may be expressed by

$$\frac{\partial \Psi}{\partial t} = \gamma^2 \left(\frac{\partial^2 \Psi}{\partial z^2} + \frac{1}{z}\frac{\partial \Psi}{\partial z} + \Psi(1-\Psi) \right), \tag{26}$$

where $z = 2r\gamma/a$. In order to describe stationary states we shall require solutions of the non-linear differential equation

$$\frac{d^2\Psi(z)}{dz^2} + \frac{1}{z}\frac{d\Psi}{dz} + \Psi(1-\Psi) = 0, \tag{27}$$

especially those that are finite at the origin, in which case

$$\Psi'(0) \equiv \frac{d}{dz}\Psi\Big|_{z=0} = 0. \tag{28}$$

The equation, not being reducible to a Painlevé transcendent, is not free from movable critical points (Valiron, 1945, pp. 293–4). The general march of solutions finite at the origin is shown in Fig. 3. Since $\Psi(1-\Psi)$ is positive if and only if $0 < \Psi < 1$, it follows from the differential equation that the turning points of all solutions in the strip $0 < \Psi < 1$ are maxima. Similarly, the only turning points for which $\Psi > 1$ or $\Psi < 0$ are minima.

The spatial distribution of population density is illustrated by AB in the case of a region of extinction, by CD in a circular region of multiplication, and by EF in the case of an annular region of multiplication.

Particular solutions of (27) subject to stated boundary conditions may be computed without much labour by the iterative method outlined in §3·9.

3·7. *The function* $Q(z;\, q)$ is defined as the solution of equation (27) subject to condition (28) and the relation $Q(0;\, q) = q$. We are concerned here with the real variable only.

If $U(z) \equiv (q - Q(z;\, q))/q(1-q)$, the differential equation satisfied by U is

$$zU'' + U' = z[1 + \lambda U - \tfrac{1}{4}(1-\lambda^2)\, U^2], \quad \text{where} \quad \lambda = 2q - 1. \tag{29}$$

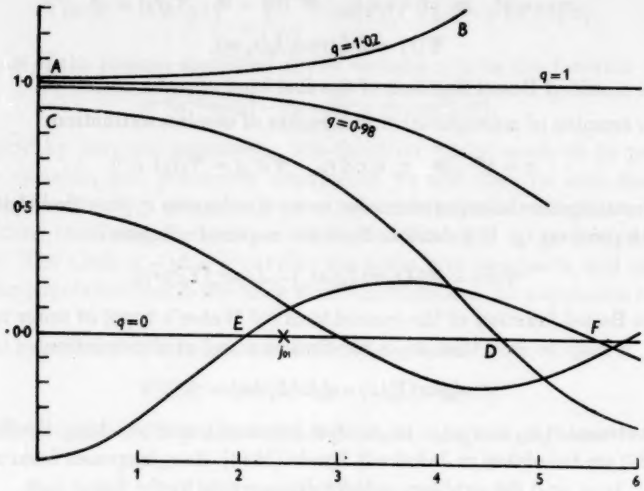


Fig. 3

Hence U tends to $1 - J_0(z)$ as $q \to 0$ and to $I_0(z) - 1$ as $q \to 1$. For computational purposes we may therefore seek an approximation to U in the immediate neighbourhood of $z = 0$ in the form

$$U = S(z) - 4q(1-q)\mathscr{C}(z),$$

where $$S(z) = \sum_1^\infty \lambda^{n-1} z^{2n}/2^2 4^2 \ldots (2n)^2,$$

may be regarded as a rough approximation

$$= [I_0(z\sqrt{\lambda}) - 1]/\lambda \quad \text{when} \quad \lambda > 0$$

$$= [1 - J_0(z\sqrt{|\lambda|})]/|\lambda| \quad \text{when} \quad \lambda < 0$$

$$= \tfrac{1}{4}z^2 \quad \text{when} \quad \lambda = 0,$$

and where $$\mathscr{C}(z) = \sum_n C_{2n} z^{2n}/2^2 4^2 \ldots (2n)^2$$

is a series of correction terms whose coefficients C_{2n} are to be determined. Only even powers are required, as may be seen either by differentiating (29) an even number of times or by considering the function $\mathscr{U}(z) \equiv U(-z)$ which is a solution of (29) with the same initial conditions. By differentiating $2n+1$ times by Leibnitz's theorem, making $z = 0$, and noting that $U_n \equiv \dfrac{d^n}{dz^n} U(z)\Big|_{z=0} = 0$ if n is odd, we obtain

$$(2n+2)\, U_{2n+2} = (2n+1)\left[\lambda U_{2n} - \tfrac{1}{4}(1-\lambda^2) \sum_{j=1}^{n-1} \binom{2n}{2j} U_{2j} U_{2n-2j}\right].$$

By substituting

$$A_{2j}/2^2 \ldots (2j)^2 \quad \text{for} \quad U_{2j}/(2j)!, \quad \text{where} \quad A_{2j} = \{\lambda^{j-1} - (1-\lambda^2)C_{2j}\},$$

we find

$$C_{2n+2} = \lambda C_{2n} + \frac{1}{4}\sum_{j=1}^{n-1}\binom{n}{j}^2 A_{2j}A_{2n-2j}.$$

Thereby we obtain in succession:

$$C_2 = C_4 = 0, \quad C_6 = 1, \quad C_8 = 11\lambda/2, \quad C_{10} = (61\lambda^2 - 16)/2,$$

$$C_{12} = \lambda(847\lambda^2 - 507)/4, \quad C_{14} = (3820\lambda^4 - 3677\lambda^2 + 488)/2,$$

$$C_{16} = \lambda(176{,}245\lambda^4 - 234{,}718\lambda^2 + 67{,}857)/8,$$

$$C_{18} = (2{,}536{,}967\lambda^6 - 4{,}317{,}594\lambda^4 + 1{,}978{,}563\lambda^2 - 162{,}816)/8.$$

Computed values of U/z^2 are given in Table 1 for equidistant values of q and of z^2. With this arrangement interpolation can be carried out reasonably safely.

Table 1. $[q - Q(z; q)]/z^2 q(1-q)$

z^2 \ q	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
0	0·2500	0·2500	0·2500	0·2500	0·2500	0·2500	0·2500	0·2500	0·2500	0·2500	0·2500
1	0·2348	0·2376	0·2405	0·2435	0·2466	0·2496	0·2528	0·2559	0·2592	0·2626	0·2661
2	0·2204	0·2255	0·2309	0·2364	0·2423	0·2483	0·2546	0·2613	0·2682	0·2754	0·2831
3	0·2068	0·2138	0·2211	0·2289	0·2373	0·2461	0·2556	0·2658	0·2767	0·2884	0·3010
4	0·1940	0·2023	0·2113	0·2211	0·2317	0·2432	0·2558	0·2696	0·2847	0·3014	0·3199
5	0·1819	0·1913	0·2016	0·2130	0·2255	0·2395	0·2551	0·2725	0·2922	0·3145	0·3399
6	0·1705	0·1806	0·1919	0·2046	0·2189	0·2351	0·2535	0·2746	0·2990	0·3275	0·3609
7	0·1597	0·1704	0·1824	0·1961	0·2118	0·2300	0·2511	0·2758	0·3052	0·3404	0·3831
8	0·1496	0·1605	0·1731	0·1876	0·2044	0·2243	0·2479	0·2761	0·3105	0·3530	0·4065
9	0·1400	0·1511	0·1639	0·1790	0·1969	0·2181	0·2439	0·2759	0·3150	0·3654	0·4312

3·8. *The stability of stationary states.* In the case of Malthusian populations with positive parameter (§§3·3 (iii), 3·5 (i)), it was apparent that the stationary state was an unstable one. Though we were able to draw valid conclusions with regard to the critical size of a habitat for a population on the point of extinction, the model is inadequate in that it does not provide a stable stationary state for a population which is safely established. The assumption of logistic growth is free from this objection.

For let Ω be a solution of (27) subject to

$$\Omega(z_a) = A, \quad \Omega(z_b) = B \quad (0 < z_a < z_b),$$

and let $\Psi(z, t)$ satisfy (26) subject to

$$\Psi(z_a, t) = A, \quad \Psi(z_b, t) = B.$$

(The proof for the special case where the conditions at $z_a = 0$ are $\Omega' = \dfrac{\partial}{\partial z}\Psi = 0$ differs only in trivial details and is omitted.)

Then from our knowledge of the physical situation we should expect $\Psi \to \Omega$ as $t \to \infty$. Provided that the necessary biological condition holds, that $\Omega(z) \geqslant 0$ in $z_a \leqslant z \leqslant z_b$ (though not zero everywhere), it can at least be shown that Ω describes a stable equilibrium.

The function $\Delta(z,t) \equiv \Psi(z,t) - \Omega(z)$ satisfies

$$\frac{1}{\gamma^2}\frac{\partial \Delta}{\partial t} = \frac{\partial^2 \Delta}{\partial z^2} + \frac{1}{z}\frac{\partial \Delta}{\partial z} + \Delta(1 - 2\Omega - \Delta). \tag{30}$$

In dealing with questions of stability we need only consider such $\Delta(z,0)$ as are infinitesimal variations of Ω. So long as Δ remains small, equation (30) remains effectively

$$\frac{1}{\gamma^2}\frac{\partial \Delta}{\partial t} = \frac{\partial^2 \Delta}{\partial z^2} + \frac{1}{z}\frac{\partial \Delta}{\partial z} + \Delta(1 - 2\Omega), \tag{31}$$

with the conditions $\Delta(z_a,t) = \Delta(z_b,t) = 0 \quad (0 \leqslant t < \infty)$.

We are therefore led to consider solutions of

$$\frac{d^2 f}{dz^2} + \frac{1}{z}\frac{df}{dz} + (\Lambda + 1 - 2\Omega) f = 0 \tag{32}$$

which satisfy $f(z_a) = f(z_b) = 0$, with $\Omega(z)$ as given. The Sturm-Liouville theorems are immediately applicable. Reduction to the standard form adopted by Titchmarsh (1946) may be brought about by the substitution $v = f\sqrt{z}$, though in this case there are advantages in following Ince (1927) and expressing (32) as

$$\frac{d}{dz}(zf') + z(\Lambda + 1 - 2\Omega) f = 0. \tag{33}$$

The values of Λ for which solutions are possible will be denoted by $\Lambda_i \; (i = 0, 1, 2, \ldots)$ and the corresponding solutions by $f_i(z)$, the suffix being the number of zeros between z_a and z_b. We now prove that Λ_0 is positive.

By multiplying (33) by Ω, substituting f_0 for f, and integrating we find

$$[z\Omega f_0']_{z_a}^{z_b} - \int_{z_a}^{z_b} z\Omega' f_0' dz + (\Lambda_0 + 1)\int_{z_a}^{z_b} z f_0 \Omega dz - 2\int_{z_a}^{z_b} z f_0 \Omega^2 dz = 0.$$

By subtracting from this a similar relationship obtained in the corresponding manner from (27) with Ω as the dependent variable, we then have

$$\Lambda_0 \int_{z_a}^{z_b} z f_0 \Omega dz = \int_{z_a}^{z_b} z f_0 \Omega^2 dz - [\Omega z f_0']_{z_a}^{z_b}. \tag{34}$$

The sign of $f_0'(z_a)$ = minus sign of $f_0'(z_b)$ = sign of both integrands, provided that Ω is positive throughout (z_a, z_b). Hence

$$\Lambda_0 \geqslant \int_{z_a}^{z_b} z f_0 \Omega^2 dz \Big/ \int_{z_a}^{z_b} z f_0 \Omega dz$$

and is positive. By Sturm's oscillation theorems, $\Lambda_0 < \Lambda_1 < \Lambda_2 \ldots$, so that all characteristic numbers are positive.

From (33) we have $$f_j \frac{d}{dz}(zf_i') + z(\Lambda_i + 1 - 2\Omega) f_j f_i = 0$$

and a similar relation with i and j interchanged. On integrating (the first term by parts) and subtracting, we find

$$(\Lambda_i - \Lambda_j)\int_{z_a}^{z_b} z f_i f_j dz = 0. \tag{35}$$

The fundamental set of solutions constitute an orthogonal system, and systems of the above kind are known to be closed (Ince, 1927). Any arbitrary function such as $\Delta(z,0)$ has the

formal development $\Delta = \sum_i A_i f_i(z)$, where the coefficients are given in the usual Fourier manner by

$$A_i = \int_{z_a}^{z_b} z\Delta(z) f_i(z)\,dz \Big/ \int_{z_a}^{z_b} z f_i^2(z)\,dz.$$

If $\Delta(z)$ is integrable over (z_a, z_b) and of bounded variation in the neighbourhood of $z = \zeta$, it may be shown that the Sturm-Liouville development given above converges to

$$\tfrac{1}{2}[\Delta(\zeta+0)+\Delta(\zeta-0)] \quad \text{at} \quad z = \zeta.$$

In problems of the type being discussed at present it is, however, permissible to assume the less general condition that $\Delta(z)$ has a continuous second derivative, and in this case the series is absolutely and uniformly convergent in (z_a, z_b) (Ince, 1927).

It now follows that the solution of (31) is

$$\Delta(z,t) = \sum_{i=0}^{\infty} A_i e^{-\Lambda_i \gamma^2 t} f_i(z).$$

As $t \to \infty$, $\Delta \to 0$, since all $\Lambda_i > 0$.

3·9. *An iterative method.* Consider, for example, the equation

$$\frac{d}{dz}(z\Psi') + zc_1(z)\Psi - zc_2(z)\Psi^2 = 0.$$

This may be written

$$\Psi(Z) = \Psi(b) + b\Psi'(b)\log(Z/b) - \int_b^Z \frac{dX}{X}\int_b^X (zc_1\Psi - zc_2\Psi^2)\,dz. \tag{36}$$

The limiting form of this equation as $b \to 0$ provided that $\Psi'(b)$ remains finite is

$$\Psi(Z) = \Psi(0) - \int_0^Z \frac{dX}{X}\int_0^X (zc_1\Psi - zc_2\Psi^2)\,dz. \tag{37}$$

If approximate values of $\Psi(z)$ are substituted on the right-hand side of (36) or (37) and the integration performed numerically, we obtain greatly improved values on the left. It is an easy matter to continue the solution step by step obtaining rough values as required by simple graphical extrapolation. Because of the remarkable rapidity of convergence it is soon found unnecessary to repeat the calculations for the earlier values (which automatically reproduce themselves).

For numerical illustration let $b = 0$, $\Psi'(0) = 0$, $\Psi(0) = 0{\cdot}5$, $c_1(z) = c_2(z) = 1$, so that the equation is

$$0{\cdot}5 - \Psi(Z) = \int_0^Z \frac{dX}{X}\int_0^X z\Psi(1-\Psi)\,dz.$$

The initial rough values are arranged in column (ii) of Table 2. Using intervals of z no smaller than $\frac{1}{2}$, and only the elementary integration formulae:

$$\int_0^h f(x)\,dx = \tfrac{1}{12}h[5f(0)+8f(h)-f(2h)],$$

$$\int_{a-h}^{a+h} f(x)\,dx = \tfrac{1}{3}h[f(a-h)+4f(a)+f(a+h)],$$

we find that a single application of the process reduces the sum of the squares of the errors to less than 1/800 its original value.

The method can be applied even if the c's are variable with finite discontinuities at a finite number of points in (b_1, b_2) provided that the interval of argument employed is not unreasonably large. Given $\Psi(b_1)$ and $\Psi'(b_1)$ new values of $\Psi(z)$ can be established progressively along (b_1, b_2). Given $\Psi(b_1)$ and $\Psi(b_2)$, however, it appears necessary to plot several integral curves radiating from the point $(b_1, \Psi(b_1))$ each with an assumed value of $\Psi'(b_1)$. After each trial this value is systematically adjusted until the condition at $z = b_2$ is satisfied.

Table 2

(i) z	(ii) approx. Ψ	(iii) $z\Psi(1-\Psi)$	(iv) $\int_0^X$ col. (iii)	(v) $\frac{1}{X}$ col. (iv)	(vi) $\int_0^Z$ col. (v)	(vii) 0·5 − (vi)	(viii) Ψ by series
0·0	0·50	0·00000	0·000000	0·000000	0·000000	0·50000	0·50000
0·5	0·48	0·12480	0·031333	0·062666	0·015711	0·48429	0·48438
1·0	0·44	0·24640	0·124267	0·124267	0·062489	0·43751	0·43761
1·5	0·36	0·34560	0·274000	0·182667	0·139445	0·36056	0·36060
2·0	0·26	0·38480	0·459867	0·229935	0·243301	0·25670	0·25681
2·5	0·13	0·28275	0·635258	0·254103	0·365530	0·13447	0·13462
3·0	0·01	0·02970	0·717450	0·239150	0·490884	0·00912	0·00924

3·10. *A more general problem.* When the vital coefficients vary from place to place in an entirely arbitrary manner the equation we have to consider is

$$\frac{\partial \Psi}{\partial t} = \tfrac{1}{4}a^2\nabla^2\Psi + c_1(x, y)\,\Psi - c_2(x, y)\,\Psi^2. \tag{38}$$

Orthodox analytical methods appear inadequate, even in one-dimensional or radially symmetrical cases. In these cases an easy extension of the argument of §3·8 shows that a sufficient but not necessary condition for the stability of solutions is that $c_2 > 0$ in the habitat considered. Because of this, simple methods or numerical solution akin to 'relaxation' were tried out for such cases, but, because of the ultimate slowness of convergence, were abandoned in favour of the method of §3·9.

3·11. *Conditions at a common boundary.* In the present treatment Ψ is regarded as continuous, and the c's as continuous except at certain points. But this is only a convenient abstraction. In nature there are discontinuities almost everywhere (see Denjoy, 1937, p. 8), and it might be thought theoretically more desirable to construct our analytical model with functions of a more general nature. The scientist, however, is not so much concerned with the behaviour of Ψ and c at a particular point as in their mean value in a small neighbourhood of that point.

Consider the situation at the point $x = b$ in a linear habitat where the 'diffusivity' is uniform throughout but where the vital coefficients in $b' \leqslant x < b$ differ from those in $b < x \leqslant b''$. (The case where one subinterval is a region of absolute extinction is excluded, and this is best treated as a limiting case with c_1 tending to $-\infty$.) An appeal to our fundamental assumptions will show that unless Ψ is the same at $b-0$ and $b+0$, the rate of population flow across the boundary would be infinitely great, proceeding in such a direction as to restore the continuity of Ψ. Similarly, unless $\partial\Psi/\partial x$ is the same at $b-0$ as at $b+0$ the rate of change of density in an infinitesimally small neighbourhood of b would be infinitely great, proceeding in such a way as to equalize the derivatives at $b-0$ and $b+0$. A similar argument holds in a radially symmetrical habitat or the more general two-dimensional one.

When employing the iterative method of §3·9 it is not necessary to define the vital coefficients at $x = b$ unless b coincides with a tabular value of the argument. In such a case it is desirable to define $c(b) = \frac{1}{2}[c(b+0)+c(b-0)]$. The continuity of c is tacitly implied in the use of formulae of numerical integration.

4. REPRODUCTIVE CAPACITY AND POPULATION DENSITY

4·1. *A law of population growth.* Strictly the logistic law is applicable only to populations which are continuous in time. It will be shown that where competition is most marked among seedlings distributed at random, a somewhat different law holds for annual plants with the analogue of the logistic law as a limiting case.

The following symbols are used:

H is a coefficient denoting the suitability of the habitat as a whole, and is resolvable into separate components. For example, $H = Wsp/\Xi$, where

W is the number of spots of ground suitable in nature and size to support one plant of the species concerned to maturity. Such a spot will be termed a cell. Cells may be isolated or adjacent in groups of two or more. In the present simplified treatment, all cells are regarded as of the same surface area, namely, one unit.

Ξ is an arbitrary area constructed to include all the cells. It is assumed that this area is insulated in the sense that it receives no seeds from without.

p denotes the proportion of seeds which actually fall in Ξ. It is assumed that the probability distribution of the seeds is even throughout Ξ.

s is the probability that a 'seeded cell' rears a seedling to maturity.

Γ denotes reproductive capacity, the number of fertile seeds produced on the average per plant.

χ denotes relative density, defined here as the ratio of actual population density to the hypothetical maximum that the habitat could support, were every cell seeded. That is, $\chi = N/Ws$, where N is the actual population number.

In the interests of completeness it might be thought desirable to introduce more ecological factors than are considered here. In most cases it will be found that the mathematical form of the final result is unchanged, particularly when the appropriate modifications are incorporated into the definitions.

The probability distribution of the number of seeds per cell will be Poissonian with parameter $N\Gamma p/\Xi = N\Gamma H/Ws = \Gamma H\chi$. The proportion of cells with at least one seed is then $1-e^{-\Gamma H\chi}$ (neglecting minor irregularities due to sampling). By reason of the definition of χ this expression is the relative density of the resulting population. Using suffixes to distinguish between the values of χ in successive generations, we then have the law

$$\chi_{n+1} = 1 - e^{-\Gamma H\chi_n}. \tag{39}$$

The ultimate stable value χ_∞ satisfies the equation

$$\log(1-\chi)+\Gamma H\chi = 0. \tag{40}$$

Typical growth 'curves' are illustrated in Fig. 4. From (40) it will be seen that as $\chi \to 1$ from lower values $\Gamma H \sim \log\frac{1}{1-\chi} \to \infty$, so that when χ is large a very considerable increase in reproductive capacity is required to bring about a perceptible increase in the population.

4·2. *Further applications.* The relation between the area 'covered' by a randomly moving insect and the distance traversed by that insect (Nicholson & Bailey, 1935) is closely analogous with the relation between area 'seeded' and the number of seeds produced by an annual plant. In many species of insect the population one year depends on the number of places suitable for oviposition encountered by the adult population of the previous year. In the simplest cases of this type, it is to be expected that relation (39) will be applicable.

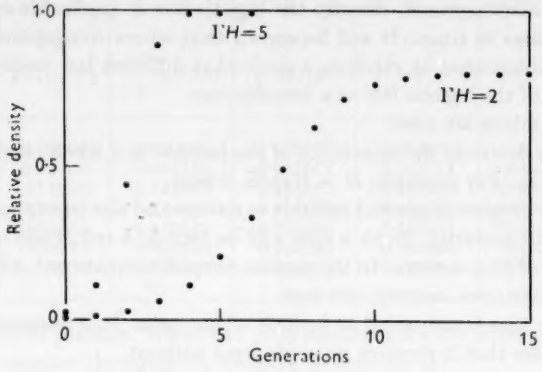


Fig. 4

Somewhat similar but more elaborate relationships have been deduced (see Jordan & Burrows, 1945; Zinsser & Wilson, 1932) in connexion with the dissemination of infectious diseases.

4·3. *Discrete logistic growth.* The recurrence relationship (39) also arises in the solution of a problem in evolutionary genetics (Skellam, 1948). Provided that $\Gamma H\chi$ is small we have as a good approximation

$$\chi_{n+1} = 1 - \frac{1 - \frac{1}{2}\Gamma H\chi_n}{1 + \frac{1}{2}\Gamma H\chi_n}. \tag{41}$$

Since $\chi_\infty = 2(\Gamma H - 1)/\Gamma H$ we may express this approximation as

$$\chi_{n+1} - \chi_n = \tfrac{1}{2}\chi_{n+1}(\chi_\infty - \chi_n),$$

corresponding to the differential equation $(7a)$. The finite difference equation is of the Ricatti type (Milne-Thomson, 1933) and has the solution

$$\chi_n = \frac{\chi_0\chi_\infty}{\chi_0 + (\chi_\infty - \chi_0)(\Gamma H)^{-n}}. \tag{42}$$

This expression is in fact the logistic law for a discrete variable. For writing $N_n/N_\infty = \chi_n/\chi_\infty$, $c = \log \Gamma H$, and denoting the independent variable by t with its origin at the centre of symmetry, we obtain the standard form $N(t) = N(\infty)/(1 + e^{-ct})$.

A simple guide to the values of ΓH for which (41) is a fair approximation to (39) is afforded by the comparison of the upper asymptotes in the two cases for the same fixed values of ΓH, as shown in Table 3.

It will be seen that χ_∞ is an increasing function of H and therefore of W, the number of suitable spots in Ξ. It therefore appears, other things being equal, that in the richer habitats

a greater proportion of the available cells are actually utilized. The inference may well be extended to perennial populations.

Table 3

$\Gamma H = 1{\cdot}0258$		1·0536	1·1157	1·277	2·0
By (40)	0·0500	0·1000	0·2000	0·400	0·797
By (41)	0·0504	0·1017	0·2074	0·433	1·000

4·4. *Two species in competition.* The classical case of two closely similar logistic species competing for the same food has been studied by Gause (1935) and Volterra (1931). The case we shall consider here is that of two closely related species of annual plant, S and S', competing in the same habitat. In order to investigate the extent to which a disadvantage in direct competition may be offset by a superiority in reproductive capacity, we shall assume the extreme condition that individual S' seedlings always fail to establish themselves in immediate competition with S seedlings (for example, the S seed may germinate earlier). All other ecological factors will be regarded as the same for both populations.

Under these conditions the S population is not adversely affected by the presence of S', and we may write

$$\Gamma H\chi + \log(1-\chi) = 0. \tag{43}$$

The number of cells unoccupied by S in the seedling stage is then $W(1-\chi)$, and these are the only ones available to S'.

It follows as in (4·2) that for an equilibrium

$$N' = sW(1-\chi)[1-e^{-N\Gamma' p/\Xi}]. \tag{44}$$

In order that the unit in terms of which we express the relative population density of S' is independent of χ, we shall denote N'/sW by χ_0'. Equation (44) then becomes

$$\chi_0' = (1-\chi)[1-e^{-\Gamma'\chi_0' H}]. \tag{45}$$

Hence, from (43) and (45),

$$\Gamma'/\Gamma = \chi \log\left(1-\frac{\chi_0'}{1-\chi}\right) \Big/ \chi_0' \log(1-\chi). \tag{46}$$

If now we allow $\chi_0' \to 0$ we find that $\Gamma'/\Gamma \to -\chi/(1-\chi)\log(1-\chi)$. This will be termed the critical value of Γ'/Γ. Unless it is exceeded S' cannot coexist with S. When S is not dense it

Table 4*a*

χ	0·0+	0·2	0·4	0·6	0·8	0·9	0·99
Critical Γ'/Γ	1·0	1·12	1·31	1·64	2·49	3·91	21·5

Table 4*b*

χ \ χ_0'	0+	0·1	0·2	0·3	0·4	0·5	0·6
0·1	1·055	1·118	1·193	1·283	1·395	1·539	1·738
0·5	1·443	1·610	1·842	2·20	2·90	∞	—

is possible for S' to survive (despite its extreme inferiority in direct competition) with only a slightly superior reproductive capacity, but when S is very dense, it is impossible for S' to survive unless its reproductive capacity is many times greater than that of S. In the case of very small populations the 'safety level' of Γ'/Γ will be somewhat greater than the critical value, because of the danger of random extinction. Table 4*a* gives the critical values of Γ'/Γ for various values of χ. Table 4*b* gives the values of Γ'/Γ required to maintain the S' population at various levels in the two cases $\chi = 0{\cdot}1$ and $0{\cdot}5$.

4·5. *Two species and two habitats.* As a simple illustration of the application of these relations consider two habitats with coefficients $H_1 = 0{\cdot}001$ and $H_2 = 0{\cdot}01$, the same for both species, and let $\Gamma = 400$ and $\Gamma' = 2000$. The seeds of S', being more numerous, might well have smaller food reserves. Whatever the cause it is assumed as before that S' seedlings do not establish themselves in *immediate* competition with S seedlings.

Since $\Gamma H_1 = 0{\cdot}4 < 1$, S cannot survive in the first habitat but does so in the second with $\chi = 0{\cdot}9803$. Hence $-\chi/(1-\chi)\log(1-\chi) = 12{\cdot}7 > \Gamma'/\Gamma = 5$. As a result S' cannot compete successfully in the second habitat though in the first it appears safely established with $\chi' = 0{\cdot}797$.

In ecological terms we could say that though the requirements of both species are the same, the species with the greater reproductive capacity is better suited to the poorer habitat, and the species with the greater ability to establish itself against competition at the seedling stage is better adapted to life in the richer habitat.

5. Biological discussion

1. The analytical model developed here assumes that dispersal is effectively at random. This is at least approximately true for large numbers of terrestrial plants and animals. The behaviour patterns of certain animals may be such, however, as to tend to lead them to more favourable conditions. To deal adequately with these cases it will be necessary to introduce a further complication into the theoretical model somewhat analogous to gravitational attraction. Nevertheless, in most instances the range of an animal's perception is small compared with its powers of dispersal, and even the more intelligent may not discriminate between two parts of a habitat differing considerably in their effect on survival. Local irregularities in the character of the environment act as stimuli initiating repeated tactic displacements, the ultimate cumulative effect of which is scarcely distinguishable from a blind randomness.

2. If a population is initially located at a centre, spreads as with a Brownian motion, and undergoes unrestricted growth, the rate of radial expansion of its contours is approximately constant (see equations 5, 5*a*). A similar law applies to the flow of a gene, but here it should be noted that γ^2 for a new advantageous mutation will almost always be very small. The oft-repeated statement that a new advantageous gene spreads through the population replacing the former wild-type allelomorph is probably true (if no more than a reasonable period of time is allotted) only in those cases where the powers of dispersal are considerable in relation to the area occupied. In assessing the roles played by various processes in bringing about speciation, it might be borne in mind that fragmentation of the distribution area is by no means a necessary condition for different accumulations of genetic diversity to arise in the same species in different parts of a comparatively uniform habitat, and ecological adaptation can always occur when the rates of gene flow are insufficient to dissipate the genetic advances being made.

3. Just as the area/volume ratio is an important concept in connexion with continuance of metabolic processes in small organisms, so is the perimeter/area concept (or some equivalent relationship) important in connexion with the survival of a community of mobile individuals. Though little is known from the study of field data concerning the laws which connect the distribution in space of the density of an annual population with its powers of dispersal, rates of growth and the habitat conditions, it is possible to conjecture the nature of this relationship in simple cases. The treatment of §3 shows that if an isolated terrestrial habitat is less than a certain critical size the population cannot survive. If the habitat is slightly greater than this the surface which expresses the density at all points is roughly dome-shaped, and for very large habitats this surface has the form of a plateau.

4. The logistic law of population growth is often compared with that of autocatalysis (D'Arcy Thompson, 1942) and the inference made that population size is limited by the available food supply. In the present paper it is shown from spatial considerations alone that many populations which are discrete in time can be expected to satisfy the logistic law—at least as a close approximation.

5. The problem of the relationship between the reproductive capacity of plants and their habitat conditions has been studied very thoroughly by Salisbury (1942), but it is by no means certain that the thesis he adopts (pp. 159, 160, 231) is entirely tenable. The analytical approach of §§4·3–4·5, though no doubt over-simplified, provides us immediately with a possible explanation of the well-known paradox that certain species flourish only in unpromising situations, and might be developed and applied with advantage to throw light on related problems in this confused and bewildering subject.

REFERENCES

BROWNLEE, J. (1911). *Proc. R. Soc. Edinb.* **31**, 262.

CHANDRASEKHAR, S. (1943). *Rev. Mod. Phys.* **15**, 1–89.

DE GEER, G. (1910). *Int. Geol. Congr.* Stockholm.

DENJOY, A. (1937). *Théorie des Fonctions de Variables Réelles*, **1**. Paris: Hermann.

FISHER, R. A. (1922). *Proc. R. Soc. Edinb.* **42**, 327.

FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

FISHER, R. A. (1937). *Ann. Eugen., Lond.*, **7**, 355–69.

GAUSE, G. F. (1935). *Vérifications experimentales de la théorie mathématique de la lutte pour la vie*. Paris: Hermann.

GODWIN, H. (1934). *New Phytol.* **33**.

HALDANE, J. B. S. (1932). *The Causes of Evolution*. Appendix. London: Longmans, Green.

INCE, E. L. (1927). *Ordinary Differential Equations*. London: Longmans, Green.

JAHNKE, E. & EMDE, F. (1945). *Tables of Functions*, 4th ed., p. 205. New York: Dover Publications.

JEANS, J. (1921). *The Dynamical Theory of Gases*. Cambridge University Press.

JORDAN, E. O. & BURROWS, W. (1945). *Textbook of Bacteriology*, 14th ed. Philadelphia: Saunders.

KENDALL, D. G. (1948). *Proc. Camb. Phil. Soc.* **44**, 591–4.

KLUYVER, J. C. (1905). *Proc. K. Akad. Vet. Amst.* 25 Oct., pp. 341–50.

KOSTITZIN, V. A. (1939). *Mathematical Biology*. London: Harrap.

LEGENDRE, A. M. (1825). *Traité des Fonctions Elliptiques*, **2**. Paris. Tables reissued with an Introduction by K. Pearson (1934). Cambridge University Press.

LOTKA, A. J. (1925). *Elements of Physical Biology*. Baltimore.

LOTKA, A. J. (1939). *Théorie Analytique des Associations Biologiques*, Part II. Paris: Hermann.

MAXWELL, J. C. (1860). *Phil. Mag.* [4], **19**, 31.

MILNE-THOMSON, L. M. (1933). *The Calculus of Finite Differences*. London: Macmillan.

NICHOLSON, A. J. & BAILEY, V. A. (1935). *Proc. Zool. Soc. Lond.*, pp. 551–98.

PEARSON, K. (1906). Mathematical Contributions to the Theory of Evolution, XV. *Drapers' Company Research Memoirs*.

RASHEVSKY, N. (1948). *Mathematical Biophysics*. University of Chicago Press.

RAYLEIGH, LORD (1919). *Phil. Mag.* [6], **37**, 321.
REID, C. (1899). *The Origin of the British Flora.* London: Dulau.
SALISBURY, E. J. (1942). *The Reproductive Capacity of Plants.* London: Bell.
SKELLAM, J. G. (1948). *Proc. Camb. Phil. Soc.* **45**, 364.
TANSLEY, A. G. (1939). *The British Islands and their Vegetation.* Cambridge University Press.
THOMPSON, D'ARCY (1942). *Growth and Form.* Cambridge University Press.
TITCHMARSH, E. C. (1946). *Eigenfunction Expansions associated with Second-Order Differential Equations.* Oxford: Clarendon Press.
ULBRICH, J. (1930). *Die Bisamratte.* Dresden: Heinrich.
USPENSKY, J. V. (1937). *Introduction to Mathematical Probability.* New York: McGraw-Hill.
VALIRON, G. (1945). *Equations Fonctionnelles.* Paris: Masson.
VOLTERRA, V. (1931). *Leçons sur la Théorie Mathématique de la Lutte pour la Vie.* Paris: Gauthier-Villars.
WATSON, G. N. (1944). *Treatise on Bessel Functions.* Cambridge University Press.
WATT, A. S. (1919). On the Causes of Failure of Natural Vegetation in British Oakwoods. *J. Ecol.* **7** (nos. 3 and 4), 173.
WRIGHT, S. (1931). *Genetics*, **16**, 97–159.
WRIGHT, W. B. (1937). *The Quaternary Ice Age.* London: Macmillan.
ZINSSER, H. & WILSON, E. B. (1932). *J. Prev. Med.* **6**, 497–514.

THE FREQUENCY DISTRIBUTION OF THE PRODUCT-MOMENT CORRELATION COEFFICIENT IN RANDOM SAMPLES OF ANY SIZE DRAWN FROM NON-NORMAL UNIVERSES

By A. K. GAYEN, *Statistical Laboratory, Cambridge*

1. Introduction

Considering the population to be specified by the normal bivariate surface, Fisher (1915) obtained the exact sampling distribution of r, the sample value of the product-moment correlation coefficient ρ. Experimental investigations were later made, by various workers, to examine how far the observed distributions of r, in samples from non-normal populations of known form, compared with the corresponding normal-theory values. E. S. Pearson (1929), in studying some experimental results for samples of 20 and 30 from two considerably non-normal populations, of which the parental correlations were respectively $\rho = 0{\cdot}5346$ and $0{\cdot}4626$, concluded that 'the results suggest that the normal bivariate surface can be mutilated and distorted to a remarkable degree without affecting the frequency distribution of r'. Other writers, following Pearson, have considered still smaller samples (of 5 to 10 in most cases) from populations varying widely in the degree of non-normality and the intensity of parental correlation. In some cases the results of their experiments may be interpreted as agreeing well with the theoretical normal-theory values but elsewhere such agreement is far from satisfactory.* The problem does not appear to have been investigated mathematically except in special cases.

Quensel (1938) derived the frequency density of r in samples from the bivariate Gram-Charlier population (alternative form, formula (18) below), but the result, as has been stated by him, is applicable only when the parental correlation ρ is zero. From the derived formulae of his study he made some observations as to the nature of the frequency distribution of r, in special cases of completely independent variates.

The present paper is devoted to a theoretical study of the distribution of non-normal r, in the general non-null case, i.e. the case in which the parental correlation ρ exists and is not necessarily small. Considering the population to be expressed by the Edgeworth form of the bivariate Type A surface (including terms as far as those in $\lambda_{30}^2, \lambda_{30}\lambda_{21}, \ldots, \lambda_{03}\lambda_{12}$ and λ_{03}^2), the mathematical form of the frequency density of r is obtained. If higher semi-invariants other than $\lambda_{40}, \lambda_{31}, \lambda_{22}, \lambda_{13}$ and λ_{04}, and $\lambda_{30}^2, \lambda_{30}\lambda_{21}, \ldots, \lambda_{03}\lambda_{12}$ and λ_{03}^2 of the population are negligible, then the derived distribution holds good for any size of sample. On the other hand, if the samples are fairly large, the formula will hold asymptotically for any form of parent population. Thus it is not unlikely that for samples of moderate size, the distribution obtained has quite an extended range of applicability, provided always the higher semi-invariants other than those considered are small. Throughout this work the values of ρ and of the λ's of the population have been assumed to be known.

The possibility of the application of Fisher's logarithmic transformation of the coefficients of correlation, namely

$$z = \tfrac{1}{2}\log_e \frac{1+r}{1-r} \quad \text{and} \quad \zeta = \tfrac{1}{2}\log_e \frac{1+\rho}{1-\rho},$$

* Cf., in particular, the results in Tables V and VI of Rider (1932). See also Editorial comments in that connexion, p. 401, footnote at end of Rider's paper.

to the derived frequency function is discussed. The asymptotic formulae for the moment constants of the distribution of z in samples from the Edgeworth population are obtainable. It is seen that for samples of moderate size, the distribution of z is still approximately normal with a mean and standard deviation depending on the fourth and the square and product of the third-order semi-invariants (being respectively the measures of 'excess' and 'skewness') of the population, apart from the size of sample N and the coefficient of correlation ρ as usual. Some practical applications of the formulae are considered.

In the course of the present work certain errors have been detected in the higher order terms of the expansions of the moment coefficients of z, given by Fisher (1921). While these terms are not important in most practical applications, they have been used by most of the workers on the present problem in examining the closeness of the approximations involved in the use of z. It is seen that the correct formulae for what David (1938, p. xxxii) calls 'Approximation I' are still closer to the actual probabilities of r.

2. Joint distribution of the three second-order central moments in samples from the bivariate Edgeworth population

Let us consider the parent population to be specified by the bivariate Edgeworth surface*

$$F(x,y) = \left\{1 + \sum_{i+j=3,4,6}\sum A_{ij}\frac{(-1)^{i+j}}{i!j!}\left(\frac{\partial}{\partial x}\right)^i\left(\frac{\partial}{\partial y}\right)^j\right\}\phi(x,y), \tag{1}$$

where
$$\phi(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)}}\exp\left[-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right]. \tag{2}$$

ρ = the population value of the correlation coefficient

$$= \lambda_{11} = \frac{\mu_{11}}{\sqrt{(\mu_{20}\mu_{02})}} = \mu_{11}, \tag{3}$$

since the variables are in standard measure, and

$$A_{30} = \lambda_{30} = \mu_{30}, \quad A_{21} = \lambda_{21} = \mu_{21}, \quad .... \tag{4}$$

$$A_{40} = \lambda_{40} = \mu_{40} - 3, \quad A_{31} = \lambda_{31} = \mu_{31} - 3\rho, \quad A_{22} = \lambda_{22} = \mu_{22} - 2\rho^2 - 1, ... \tag{5}$$

$$A_{60} = 10\lambda_{30}^2, \quad A_{51} = 10\lambda_{30}\lambda_{21}, \quad A_{42} = 4\lambda_{30}\lambda_{12} + 6\lambda_{21}^2, \quad A_{33} = \lambda_{30}\lambda_{03} + 9\lambda_{21}\lambda_{12}, ... \tag{6}$$

Here the $\lambda_{ij}\left(= \dfrac{\kappa_{ij}}{\kappa_{20}^{\frac{1}{2}i}\kappa_{02}^{\frac{1}{2}j}} = \kappa_{ij}, \text{ since } \kappa_{20} = \kappa_{02} = 1\right)$ will be termed the semi-invariants of the population. Using Quensel's (1938, pp. 76–80) results, the characteristic function of the joint distribution of the three second-order central moments,

$$m_{20} = \frac{\Sigma(x-\bar{x})^2}{N}, \quad m_{11} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{N} \quad \text{and} \quad m_{02} = \frac{\Sigma(y-\bar{y})^2}{N}, \tag{7}$$

where
$$\bar{x} = \frac{\Sigma(x)}{N}, \quad \bar{y} = \frac{\Sigma(y)}{N}, \tag{8}$$

* Extending his generalized law of error to two dimensions, Edgeworth considers terms involving moments up to the third order only, $i+j=3$. The form of approximation (1), where terms in

$$\lambda_{30}, \lambda_{21}, ..., \lambda_{03}; \lambda_{40}, \lambda_{31}, ..., \lambda_{04} \quad \text{and} \quad \lambda_{30}^2, \lambda_{30}\lambda_{21}, ..., \lambda_{03}^2$$

are included, has been considered by Wicksell (1917), who has studied also, among other things, the regression of its characteristics. In the present work it will be referred to as the Edgeworth surface or population.

in random samples of size N, from the population (1), is found to be

$$U(t_{20}, t_{11}, t_{02}) = E\{\exp[it_{20}m_{20} + it_{11}m_{11} + it_{02}m_{02}]\}$$

$$= D^{-\frac{1}{2}(N-1)}\Big[1 + \frac{(N-1)^2}{8N(1-\rho^2)^2}\{\lambda_{40}\tau_{20}^2 + 4\lambda_{31}\tau_{20}\tau_{11} + 2\lambda_{22}(\tau_{20}\tau_{02} + 2\tau_{11}^2) + 4\lambda_{13}\tau_{02}\tau_{11} + \lambda_{04}\tau_{02}^2\}$$

$$+ \frac{(N-1)(N-2)}{12N(1-\rho^2)^3}\{\lambda_{30}^2\tau_{20}^3 + 6\lambda_{30}\lambda_{21}\tau_{20}^2\tau_{11} + 6\lambda_{30}\lambda_{12}\tau_{20}\tau_{11}^2 + 3\lambda_{21}^2(\tau_{20}^2\tau_{02} + 2\tau_{20}\tau_{11}^2)$$

$$+ 2\lambda_{30}\lambda_{03}\tau_{11}^3 + 6\lambda_{21}\lambda_{12}(\tau_{11}^3 + 2\tau_{20}\tau_{11}\tau_{02}) + 3\lambda_{12}^2(\tau_{20}\tau_{02}^2 + 2\tau_{02}\tau_{11}^2)$$

$$+ 6\lambda_{03}\lambda_{21}\tau_{02}\tau_{11}^2 + 6\lambda_{03}\lambda_{12}\tau_{02}^2\tau_{11} + \lambda_{03}^2\tau_{02}^3\}\Big], \quad (9)$$

where $i = \sqrt{(-1)}$, and

$$D = \left\{\left(1 - 2\frac{it_{20}}{N}\right)\left(1 - 2\frac{it_{02}}{N}\right) - 4\rho^2\frac{it_{20}}{N}\frac{it_{02}}{N} - 2\rho\frac{it_{11}}{N} - (1-\rho^2)\left(\frac{it_{11}}{N}\right)^2\right\}, \quad (10)$$

and

$$\left.\begin{aligned}\tau_{20} &= \frac{1}{D} - \frac{2}{ND}(\rho^2 it_{20} + \rho it_{11} + it_{02}) - 1,\\ \tau_{11} &= \rho - \frac{\rho}{D} + \frac{1}{ND}(2\rho it_{20} + (1+\rho^2)it_{11} + 2\rho it_{02}),\\ \tau_{02} &= \frac{1}{D} - \frac{2}{ND}(it_{20} + \rho it_{11} + \rho^2 it_{02}) - 1.\end{aligned}\right\} \quad (11)$$

As a suitable check to the result (9), the mean and the variance of the distribution of the product-moment coefficient m_{11}, calculated directly from this formula, were found to agree with Pepper's (1929, pp. 233–4) results, namely

and

$$\left.\begin{aligned}\text{mean } m_{11} &= \frac{(N-1)}{N}\rho,\\ \text{var } m_{11} &= \frac{(N-1)}{N^3}\{1 + (N-1)\mu_{22} - (N-2)\rho^2\}\\ &= \frac{(N-1)}{N^3}\{N(1+\rho^2) + (N-1)\lambda_{22}\},\end{aligned}\right\} \quad (12)$$

which he deduced by a different method.

Now to derive the expression for the joint frequency function of m_{20}, m_{11} and m_{02}, we shall have to evaluate the integral*

$$f(m_{20}, m_{11}, m_{02}) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int\int U(t_{20}, t_{11}, t_{02})\exp[-it_{20}m_{20} - it_{11}m_{11} - it_{02}m_{02}]\,dt_{20}dt_{11}dt_{02}, \quad (13)$$

a typical term of which is of the form

$$f_{\nu;\nu_1,\nu_2,\nu_3}(m_{20}, m_{11}, m_{02}) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int\int D^{-\frac{1}{2}(N-1)-\nu}(it_{20})^{\nu_1}(it_{11})^{\nu_2}(it_{02})^{\nu_3}$$

$$\times \exp[-it_{20}m_{20} - it_{11}m_{11} - it_{02}m_{02}]\,dt_{20}dt_{11}dt_{02} \quad (14)$$

for ν varying from 0 to 3 and $\nu_1 + \nu_2 + \nu_3 \leqslant 3$.

The value of (14) for $\nu_1 = \nu_2 = \nu_3 = 0$, i.e. the integral

$$f_{\nu;0,0,0}(m_{20}, m_{11}, m_{02}) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int\int D^{-\frac{1}{2}(N-1)-\nu}\exp[-it_{20}m_{20} - it_{11}m_{11} - it_{02}m_{02}]\,dt_{20}dt_{11}dt_{02} \quad (15)$$

* f is defined as zero except for $m_{20} > 0$, $m_{20}m_{02} > m_{11}^2$.

will be found from Romanovsky's (1925) formula (111), as

$$f_\nu(m_{20}, m_{11}, m_{02}) = \frac{1}{4\pi}\left(\frac{N^2}{1-\rho^2}\right)^{\frac{1}{2}(N-1)+\nu} \frac{(m_{20}m_{02}-m_{11}^2)^{\frac{1}{2}(N-4)+\nu}}{(N+2\nu-3)!} \times \exp\left[-\frac{N}{2(1-\rho^2)}(m_{20}-2\rho m_{11}+m_{02})\right]. \quad (16)^*$$

Accordingly, whatever may be the value of ν_1, ν_2 and ν_3, the integral in (14) may be evaluated by successive differentiation of (16), with respect to m_{20}, m_{11} and m_{02}. Thus we have

$$f_{\nu;\nu_1,\nu_2,\nu_3}(m_{20}, m_{11}, m_{02}) = (-1)^{\nu_1+\nu_2+\nu_3}\left(\frac{\partial}{\partial m_{20}}\right)^{\nu_1}\left(\frac{\partial}{\partial m_{11}}\right)^{\nu_3}\left(\frac{\partial}{\partial m_{02}}\right)^{\nu_2} f_\nu(m_{20}, m_{11}, m_{02}). \quad (17)$$

The formulae (102)–(104) of Quensel, giving the forms of a typical term in the joint distribution of m_{20}, m_{11} and m_{02}, in samples from the Charlier Type A population of which the correlation ρ is not necessarily small, involve functions of the type (17). But excepting in the special case $\rho = 0$, they have not been evaluated in explicit form. Quensel derives the joint distribution of m_{20}, m_{11} and m_{02}, and the frequency density of r, for the alternative form of expansion of the Type A population, namely,

$$F(x,y) = \left\{1 + \sum\sum_{i+j=1 \text{ and } \geqslant 3} A'_{ij}\frac{(-1)^{i+j}}{i!\,j!}\left(\frac{\partial}{\partial x}\right)^i\left(\frac{\partial}{\partial y}\right)^j\right\}\phi(x)\,\phi(y), \quad (18)$$

where

$$\phi(x) = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2} \quad \text{and} \quad \phi(y) = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}y^2}, \quad (19)$$

a form applicable only when ρ is small. The advantage of such a form of frequency surface, (18), is that the characteristic function of the joint distribution of the three second-order central moments comes out to be very simple, for the terms containing ρ do not appear in it. As may be noticed, the expressions (10) and (11) occurring in the characteristic function (9), are considerably reduced in form when ρ is zero.

In the present case, as we need to consider only values of ν_1, ν_2 and ν_3 up to $\nu_1+\nu_2+\nu_3 \leqslant 3$, the complete expressions for (17) can be deduced for any ρ without much difficulty. Substituting them in (13), and after some simplification, if we write

$$\left.\begin{aligned}(1-\rho^2)\,M_{20} &= m_{20}-2\rho m_{11}+\rho^2 m_{02},\\ (1-\rho^2)\,M_{11} &= \rho m_{20}-(1+\rho^2)\,m_{11}+\rho m_{02},\\ (1-\rho^2)\,M_{02} &= \rho^2 m_{20}-2\rho m_{11}+m_{02},\end{aligned}\right\} \quad (20)$$

then the joint distribution of m_{20}, m_{11}, and m_{02} will be found as

$$\begin{aligned}f(m_{20}, m_{11}, m_{02}) &= f_0(m_{20}, m_{11}, m_{02})\Big[1 + \frac{(N-1)}{8N(N+1)(1-\rho^2)^2}\{\lambda_{40}[N^2M_{20}^2 - 2N(N+1)\,M_{20} + (N+1)(N-1)]\\ &\quad - 4\lambda_{31}[N^2M_{20}M_{11} - N(N+1)(\rho M_{20}+M_{11}) + (N+1)(N-1)\rho]\\ &\quad + 2\lambda_{22}[N^2(M_{20}M_{02}+2M_{11}^2) - N(N+1)(M_{20}+4\rho M_{11}+M_{02}) + (N+1)(N-1)(1+2\rho^2)]\\ &\quad - 4\lambda_{13}[N^2M_{02}M_{11} - N(N+1)(\rho M_{02}+M_{11}) + (N+1)(N-1)\rho]\\ &\quad + \lambda_{04}[N^2M_{02}^2 - 2N(N+1)\,M_{02} + (N+1)(N-1)]\}\end{aligned}$$

* Quensel's formulae (100)–(101), giving the expression for the function $f_\nu(m_{20}, m_{11}, m_{02})$, involve a printer's error, as the correct form of C_n should be (in his notation)

$$C_0\left(\frac{N^2}{1-r^2}\right)^n \frac{(N-3)!}{(N+2n-3)!} \quad \text{instead of} \quad C_0\left(\frac{N}{1-r^2}\right)^{2n} \frac{(N-3)!}{(N+2n-3)!}$$

$$+\frac{(N-2)}{12N(N+1)(N+3)(1-\rho^2)^3}\Big\{\lambda_{30}^2[N^3M_{20}^3-3N^2(N+3)M_{20}^2+3N(N+3)(N+1)M_{20}$$
$$-(N+3)(N+1)(N-1)]$$
$$-6\lambda_{30}\lambda_{21}[N^3M_{20}^2M_{11}-N^2(N+3)M_{20}(\rho M_{20}+2M_{11})$$
$$+N(N+3)(N+1)(2\rho M_{20}+M_{11})-(N+3)(N+1)(N-1)\rho]$$
$$+6\lambda_{30}\lambda_{12}[N^3M_{20}M_{11}^2-N^2(N+3)M_{11}(2\rho M_{20}+M_{11})+N(N+3)(N+1)\rho(\rho M_{20}+2M_{11})$$
$$-(N+3)(N+1)(N-1)\rho^2-\frac{N^2}{(N-2)}(m_{20}m_{02}-m_{11}^2)(NM_{20}-(N+3))]$$
$$+3\lambda_{21}^2[N^3M_{20}(M_{20}M_{02}+2M_{11}^2)-N^2(N+3)\{(M_{20}+2M_{02})M_{20}+2(2\rho M_{20}+M_{11})M_{11}\}$$
$$+N(N+3)(N+1)\{2(1+\rho^2)M_{20}+4\rho M_{11}+M_{02}\}$$
$$-(N+3)(N+1)(N-1)(1+2\rho^2)+\frac{2N^2}{(N-2)}(m_{20}m_{02}-m_{11}^2)(NM_{20}-(N+3))]$$
$$-2\lambda_{30}\lambda_{03}[N^3M_{11}^3-3N^2(N+3)\rho M_{11}^2+3N(N+3)(N+1)\rho^2M_{11}$$
$$-(N+3)(N+1)(N-1)\rho^3-\frac{3N^2}{(N-2)}(m_{20}m_{02}-m_{11}^2)(NM_{11}-(N+3)\rho)]$$
$$-6\lambda_{21}\lambda_{12}[N^3M_{11}(M_{11}^2+2M_{20}M_{02})-N^2(N+3)\{2M_{20}(\rho M_{02}+M_{11})+M_{11}(2M_{02}+3\rho M_{11})\}$$
$$+N(N+3)(N+1)\{2\rho(M_{20}+M_{02})+(2+3\rho^2)M_{11}\}$$
$$-(N+3)(N+1)(N-1)(2+\rho^2)\rho+\frac{N^2}{(N-2)}(m_{20}m_{02}-m_{11}^2)(NM_{11}-(N+3)\rho)]$$
$$+3\lambda_{12}^2[N^3M_{02}(M_{20}M_{02}+2M_{11}^2)-N^2(N+3)\{(M_{02}+2M_{20})M_{02}+2(2\rho M_{02}+M_{11})M_{11}\}$$
$$+N(N+3)(N+1)\{2(1+\rho^2)M_{02}+4\rho M_{11}+M_{20}\}$$
$$-(N+3)(N+1)(N-1)(1+2\rho^2)+\frac{2N^2}{(N-2)}(m_{20}m_{02}-m_{11}^2)(NM_{02}-(N+3))]$$
$$+6\lambda_{03}\lambda_{21}[N^3M_{02}M_{11}^2-N^2(N+3)M_{11}(2\rho M_{02}+M_{11})+N(N+3)(N+1)\rho(\rho M_{02}+2M_{11})$$
$$-(N+3)(N+1)(N-1)\rho^2-\frac{N^2}{(N-2)}(m_{20}m_{02}-m_{11}^2)(NM_{02}-(N+3))]$$
$$-6\lambda_{03}\lambda_{12}[N^3M_{02}^2M_{11}-N^2(N+3)M_{02}(\rho M_{02}+2M_{11})$$
$$+N(N+3)(N+1)(2\rho M_{20}+M_{11})-(N+3)(N+1)(N-1)\rho]$$
$$+\lambda_{03}^2[N^3M_{02}^3-3N^2(N+3)M_{02}^2+3N(N+3)(N+1)M_{02}-(N+3)(N+1)(N-1)]\Big\}\Big], \quad (21)$$

where

$$f_0(m_{20},m_{11},m_{02})=\frac{1}{4\pi}\left(\frac{N^2}{1-\rho^2}\right)^{\frac{1}{4}(N-1)}\frac{(m_{20}m_{02}-m_{11}^2)^{\frac{1}{4}(N-4)}}{\Gamma(N-2)}\exp\left[-\frac{N}{2(1-\rho^2)}(m_{20}-2\rho m_{11}+m_{02})\right] \quad (22)$$

is the normal-theory frequency function of m_{20}, m_{11} and m_{02}.

3. The distribution of the product-moment correlation coefficient

In the joint distribution (21) of the variables m_{20}, m_{11} and m_{02} let us introduce the new variable r, the product-moment correlation coefficient of the sample, by the substitution

$$m_{11} = r\sqrt{(m_{20}m_{02})}. \tag{23}$$

Next, as usual, putting $\quad m_{20} = we^t, \quad m_{02} = we^{-t}$ and $r = r,$ (24)

integrating for w and writing $\quad I_{(N)} = \int_0^\infty \frac{dt}{(\cosh t - \rho r)^N},$ (25)

we obtain the frequency density of r, in the form

$$f(r) = \frac{N-2}{\pi}(1-\rho^2)^{\frac{1}{2}(N-1)}(1-r^2)^{\frac{1}{2}(N-4)}[I_{(N-1)} + \{(\nu_{40})I_{(N-1)} - r(\nu_{31})I_{(N)} + (r^2(\nu_{13}) - (\nu_{04}))I_{(N+1)}\}$$
$$+ \{(\nu_{60})I_{(N-1)} - r(\nu_{51})I_{(N)} + (r^2(\nu_{42}) - (\nu_{24}))I_{(N+1)} - (r^3(\nu_{15}) - r(\nu_{06}))I_{(N+2)}\}], \tag{26}$$

where (ν_{ij}) are expressions involving, besides N and ρ, the fourth and the square and product of the third order semi-invariants of the population, for $i+j = 4$ and 6 respectively.

The result (26) undergoes further simplification when the recurrence relations (Co-operative Study (1917) formula xxxviii) namely,

$$(N-1)I_{(N-1)} + (2N-1)\rho r I_{(N)} = N(1-\rho^2 r^2)I_{(N+1)}, \tag{27}$$

between the integrals of the form (25), are used. Also putting

$$\psi(r,\rho) = \frac{N-2}{\pi}(1-\rho^2)^{\frac{1}{2}(N-1)}(1-r^2)^{\frac{1}{2}(N-4)}I_{(N-1)} \tag{28}$$

and differentiating successively with respect to ρ, it will be seen that

$$(N-1)\rho\psi(r,\rho) + (1-\rho^2)\frac{\partial}{\partial\rho}\psi(r,\rho) = \frac{(N-2)(N-1)}{\pi}(1-\rho^2)^{\frac{1}{2}(N+1)}(1-r^2)^{\frac{1}{2}(N-4)}rI_{(N)}, \tag{29}$$

$$(N-1)(N\rho^2+1)\psi(r,\rho) + 2(N-1)\rho(1-\rho^2)\frac{\partial}{\partial\rho}\psi(r,\rho) + (1-\rho^2)^2\frac{\partial^2}{\partial\rho^2}\psi(r,\rho)$$
$$= \frac{(N-2)}{\pi}(N-1)N(1-\rho^2)^{\frac{1}{2}(N+3)}(1-r^2)^{\frac{1}{2}(N-4)}r^2I_{(N+1)}, \tag{30}$$

and $\quad (N+1)(N-1)(N\rho^2+3)\rho\psi(r,\rho) + 3(N-1)(N\rho^2+1)(1-\rho^2)\frac{\partial}{\partial\rho}\psi(r,\rho)$

$$+ 3(N-1)\rho(1-\rho^2)^2\frac{\partial^2}{\partial\rho^2}\psi(r,\rho) + (1-\rho^2)^3\frac{\partial^3}{\partial\rho^3}\psi(r,\rho)$$
$$= \frac{(N-2)}{\pi}(N-1)N(N+1)(1-\rho^2)^{\frac{1}{2}(N+5)}(1-r^2)^{\frac{1}{2}(N-4)}r^3I_{(N+2)}. \tag{31}$$

Using (27) in (26), and then applying the results of (28)–(31), we obtain the required frequency density of r as

$$f(r) = \psi(r,\rho) + \frac{(N-1)}{8N(N+1)}\left\{L_{4,1}\frac{\partial}{\partial\rho}\psi(r,\rho) + L_{4,2}\frac{\partial^2}{\partial\rho^2}\psi(r,\rho)\right\}$$
$$+ \frac{(N-2)}{12N(N+1)(N+3)}\left\{L_{6,1}\frac{\partial}{\partial\rho}\psi(r,\rho) + L_{6,2}\frac{\partial^2}{\partial\rho^2}\psi(r,\rho) + L_{6,3}\frac{\partial^3}{\partial\rho^3}\psi(r,\rho)\right\}, \tag{32}$$

where $\psi(r,\rho)$, given in (28), is the normal theory frequency function of r, in samples of size N. Further

$$\left.\begin{aligned} L_{4,1} &= 3\rho(\lambda_{40}+\lambda_{04})-4(\lambda_{31}+\lambda_{13})+2\rho\lambda_{22}, \\ L_{4,2} &= \rho^2(\lambda_{40}+\lambda_{04})-4\rho(\lambda_{31}+\lambda_{13})+2(2+\rho^2)\lambda_{22}, \end{aligned}\right\} \tag{33}$$

these being expressions involving ρ and the fourth order semi-invariants of the population, while

$$\left.\begin{aligned} L_{6,1} &= -15\rho(\lambda_{30}^2+\lambda_{03}^2)-9\rho\left(1+\frac{2}{N-2}\right)(\lambda_{21}^2+\lambda_{12}^2)-\frac{6}{N-2}\lambda_{30}\lambda_{03} \\ &\quad +6\left(2+\frac{1}{N-2}\right)\lambda_{21}\lambda_{12}+18(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+\frac{18\rho}{N-2}(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21}), \\ L_{6,2} &= -9\rho^2(\lambda_{30}^2+\lambda_{03}^2)-3\left(4+5\rho^2-\frac{2(2-5\rho^2)}{N-2}\right)(\lambda_{21}^2+\lambda_{12}^2)-\frac{18\rho}{N-2}\lambda_{30}\lambda_{03} \\ &\quad +18\rho\left(2+\frac{1}{N-2}\right)\lambda_{21}\lambda_{12}+30\rho(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})-6\left(2+\frac{2-5\rho^2}{N-2}\right)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21}), \\ L_{6,3} &= -\rho^3(\lambda_{30}^2+\lambda_{03}^2)-3\rho\left(2+\rho^2-\frac{2(1-\rho^2)}{N-2}\right)(\lambda_{21}^2+\lambda_{12}^2)+2\left(1+\frac{3(1-\rho^2)}{N-2}\right)\lambda_{30}\lambda_{03} \\ &\quad +6\left(1+2\rho^2-\frac{1-\rho^2}{N-2}\right)\lambda_{21}\lambda_{12}+6\rho^2(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12}) \\ &\quad -6\rho\left(1+\frac{1-\rho^2}{N-2}\right)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21}), \end{aligned}\right\} \tag{34}$$

these being expressions involving N, ρ and the square and product of the third-order semi-invariants of the population.

Of the two subscripts in each L coefficient, the first one refers to the order of the semi-invariants occurring in it and the second to that of the derivative of $\psi(r,\rho)$, to which it is attached.

As a check to (32), it will be seen that

$$\int_{-1}^{1} f(r)\,dr = \int_{-1}^{1} \psi(r,\rho)\,dr = 1, \tag{35}$$

since

$$\int_{-1}^{1}\left(\frac{\partial}{\partial\rho}\right)^{\nu}\psi(r,\rho)\,dr = \left(\frac{\partial}{\partial\rho}\right)^{\nu}\left[\int_{-1}^{1}\psi(r,\rho)\,dr\right] = 0. \tag{36}$$

Now that the frequency distribution of non-normal r is found, as in (32), in terms of the derivatives with respect to ρ of the normal-theory frequency function of r, the moments of any order of the former may be obtained without difficulty. For

$$\int_{-1}^{1} r^k\left(\frac{\partial}{\partial\rho}\right)^{\nu}\psi(r,\rho)\,dr = \left(\frac{\partial}{\partial\rho}\right)^{\nu}\left[\int_{-1}^{1} r^k\psi(r,\rho)\,dr\right] = \left(\frac{\partial}{\partial\rho}\right)^{\nu}\mu_k'(r), \tag{37}$$

where $\mu_k'(r)$ is the kth moment of the normal-theory distribution of r, and is already known (Co-operative Study (1917) and Romanovsky (1925)).

Up to N^{-2} the mean and variance of r, calculated directly from (32), are found to be

$$\begin{aligned} \text{mean } r &= \rho-\frac{1}{N}\{\tfrac{1}{2}\rho(1-\rho^2)-(\tfrac{3}{8}\rho(\lambda_{40}+\lambda_{04})-\tfrac{1}{2}(\lambda_{31}+\lambda_{13})+\tfrac{1}{4}\rho\lambda_{22})\} \\ &\quad -\frac{1}{N^2}\{\tfrac{3}{8}\rho(1-\rho^2)(1+3\rho^2) \\ &\qquad +\tfrac{15}{16}\rho(1-\rho^2)(\lambda_{40}+\lambda_{04})+\tfrac{1}{4}(9\rho^2-5)(\lambda_{31}+\lambda_{13})-\tfrac{1}{8}\rho(9\rho^2+7)\lambda_{22} \\ &\qquad +\tfrac{5}{4}\rho(\lambda_{30}^2+\lambda_{03}^2)+\tfrac{3}{4}\rho(\lambda_{21}^2+\lambda_{12}^2)-\tfrac{3}{2}(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})-\lambda_{21}\lambda_{12}\}, \end{aligned} \tag{38}$$

and

$$\operatorname{var} r = \frac{1}{N}\{(1-\rho^2)^2 + \tfrac{1}{4}\rho^2(\lambda_{40}+\lambda_{04}) - \rho(\lambda_{31}+\lambda_{13}) + \tfrac{1}{2}(2+\rho^2)\lambda_{22}\}$$
$$+ \frac{1}{N^2}\{\tfrac{1}{2}(1-\rho^2)(2+11\rho^2)$$
$$- (\tfrac{1}{4}\rho^2(11-15\rho^2)(\lambda_{40}+\lambda_{04}) - \rho(7-11\rho^2)(\lambda_{31}+\lambda_{13}) + \tfrac{1}{2}(10-11\rho^2-11\rho^4)\lambda_{22})$$
$$- (\tfrac{3}{2}\rho^2(\lambda_{30}^2+\lambda_{03}^2) + \tfrac{1}{2}(4+5\rho^2)(\lambda_{21}^2+\lambda_{12}^2) - 5\rho(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})$$
$$+ 2(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21}) - 6\rho\lambda_{21}\lambda_{12})\}. \quad (39)$$

It will be seen that to N^{-1} they are in agreement with the known results.

From a different approach, Tschuprow (1925, pp. 114–15, English Translation), reached an asymptotic expression, correct to N^{-2}, for the mean value of the sampling distribution of r, which, as has been stated, holds good for any law of dependence. The formula involves the parameters

$$q_{ij} = \frac{\mu_{ij}}{\mu_{20}^{\frac{1}{2}i}\mu_{02}^{\frac{1}{2}j}} \quad (40)$$

of the parent population, up to $i+j=6$ (Tschuprow's notation for q_{ij} was $r_{ij/}$, but the former has been used here to avoid confusion; also his $r_{1/1}$ is our ρ).

As a very useful check the result obtained from his formula by the substitution of

$$\left.\begin{array}{llll} q_{30}=\lambda_{30} & & q_{40}=\lambda_{40}+3, & \\ q_{21}=\lambda_{21} & & q_{31}=\lambda_{31}+3\rho, & \\ \ldots & \ldots & q_{22}=\lambda_{22}+2\rho^2+1, & \\ & & \ldots \quad \ldots & \ldots, \end{array}\right\} \quad (41)$$

and

$$\left.\begin{array}{l} q_{60}=10\lambda_{30}^2+15\lambda_{40}+15, \\ q_{51}=10\lambda_{30}\lambda_{21}+10\lambda_{31}+5\rho\lambda_{40}+15\rho, \\ q_{42}=4\lambda_{30}\lambda_{12}+6\lambda_{21}^2+6\lambda_{22}+8\rho\lambda_{31}+\lambda_{40}+3(1+4\rho^2), \\ q_{33}=\lambda_{30}\lambda_{03}+9\lambda_{21}\lambda_{12}+3(\lambda_{31}+\lambda_{13})+9\rho\lambda_{22}+3\rho(3+2\rho^2), \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \end{array}\right\} \quad (42)$$

(which are the values of the above q-parameters, equation (40), for the Edgeworth population (1)), was found to agree with our expression (38), for mean r, up to the same order of approximation.

As a special case, for $\rho = 0$, the frequency function of r, equation (32), is found to be

$$f_0(r) = \psi(r,0) + \frac{(N-1)}{2N(N+1)}\left\{l_{4,1}\left[\frac{\partial}{\partial\rho}\psi(r,\rho)\right]_{\rho=0} + l_{4,2}\left[\frac{\partial^2}{\partial\rho^2}\psi(r,\rho)\right]_{\rho=0}\right\}$$
$$+ \frac{(N-2)}{6N(N+1)(N+3)}\left\{l_{6,1}\left[\frac{\partial}{\partial\rho}\psi(r,\rho)\right]_{\rho=0} + l_{6,2}\left[\frac{\partial^2}{\partial\rho^2}\psi(r,\rho)\right]_{\rho=0} + l_{6,3}\left[\frac{\partial^3}{\partial\rho^3}\psi(r,\rho)\right]_{\rho=0}\right\}, \quad (43)$$

where

$$l_{4,1} = -(\lambda_{31}+\lambda_{13}), \quad l_{4,2} = \lambda_{22}, \quad (44)$$

and

$$\left.\begin{array}{l} l_{6,1} = 9(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12}) - \dfrac{3}{N-2}\lambda_{30}\lambda_{03} + \dfrac{3(2N-3)}{N-2}\lambda_{21}\lambda_{12}, \\ l_{6,2} = -6\dfrac{N-3}{N-2}(\lambda_{21}^2+\lambda_{12}^2) - 6\dfrac{N-1}{N-2}(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21}), \\ l_{6,3} = \dfrac{N+1}{N-2}\lambda_{30}\lambda_{03} + 3\dfrac{N-3}{N-2}\lambda_{21}\lambda_{12}. \end{array}\right\} \quad (45)$$

On simplification, the result (43) has been put in Quensel's form (his formula (125), p. 100), namely

$$f_0(r) = g_0(r) - E_1 \frac{\partial}{\partial r} g_1(r) + \frac{1}{2!} E_2 \frac{\partial^2}{\partial r^2} g_2(r) - \frac{1}{3!} E_3 \frac{\partial^3}{\partial r^3} g_3(r), \tag{46}$$

where

$$g_\nu(r) = \frac{1}{B(\frac{1}{2}, \frac{1}{2}(N-2)+\nu)} (1-r^2)^{\frac{1}{2}(N-4)+\nu}, \tag{47}$$

$$E_1 = \frac{1}{2(N+1)} \left(\frac{2}{N} \frac{\Gamma^2(\frac{1}{2}N)}{\Gamma^2 \frac{1}{2}(N-1)} \right) \left\{ -(\lambda_{31}+\lambda_{13}) + \frac{N-2}{N-1} \left(\frac{3}{N+3} (\lambda_{30}\lambda_{21} + \lambda_{03}\lambda_{12}) + \frac{2}{N+1} \lambda_{21}\lambda_{12} \right) \right\}, \tag{48}$$

$$E_2 = \frac{(N-1)(N-2)}{N(N+1)^2} \left\{ \lambda_{22} - \frac{2}{N+3} \left(\lambda_{30}\lambda_{12} + \lambda_{03}\lambda_{21} + \frac{N-3}{N-1} (\lambda_{21}^2 + \lambda_{12}^2) \right) \right\}, \tag{49}$$

and

$$E_3 = \frac{N^2(N-2)}{(N-1)(N+1)^2(N+3)^2} \left(\frac{2}{N} \frac{\Gamma^2(\frac{1}{2}N)}{\Gamma^2 \frac{1}{2}(N-1)} \right) \left\{ \lambda_{20}\lambda_{03} + \frac{3(N-3)}{N+1} \lambda_{21}\lambda_{12} \right\}. \tag{50}$$

The E-coefficients, formulae (48)–(50), will be found to agree with the corresponding results deduced from Quensel's expression (126), by putting ρ or $\lambda_{11} = 0$.

The simplicity of the frequency function of r obtained in (32) suggests that it might have been reached, otherwise, by fewer steps, had it been possible to express the joint distribution of m_{20}, m_{11} and m_{02}, or its characteristic function or the population itself, in terms of the derivatives with respect to ρ of the corresponding normal theory functions. But attempts on these lines have not been successful. Obviously, the population cannot be expressed in terms of the derivatives with respect to ρ of the normal bivariate surface. It will be noticed, however, that the use of recurrence relation (27) has contributed a good deal towards a finally compact form, as in (32), of the frequency function of r. Neither in the case of the joint distribution of the quadratic moments nor in the case of its characteristic function has it been possible to make use of similar recurrence relations.

4. The asymptotic properties of the derived distribution and the corrective functions for population excess and skewness

It will be noted that the distribution of r, obtained in (32), holds good for any size of sample from the population (1), if the values of the higher λ's of the population, other than $\lambda_{40}, \lambda_{31}, \ldots, \lambda_{04}$ and $\lambda_{30}^2, \lambda_{30}\lambda_{21}, \ldots, \lambda_{03}^2$ are negligible. Tschuprow's formulae for the moments of r indicate that the higher population parameters can only occur in terms in N^{-2} (or in higher negative powers of N). So that, whatever may be the population form, provided the samples are fairly large, the same formula (32) will be sufficiently accurate for the distribution of r. Thus it is not unlikely that the derived law has quite an extended range of applicability, in moderate size of samples, provided always the fifth and higher order semi-invariants of the population are small.

As is seen, the expression for $f(r)$, (32), consists of three components. The first being the normal-theory frequency function of r, the other two may be called the corrective functions due to population excess (as measured by the fourth-order semi-invariants) and skewness (as measured by the square and product of the third-order semi-invariants) respectively.

Let us write

$$\left.\begin{aligned} \psi_{4,1}(r,\rho) &= \frac{(N-1)}{8N(N+1)} \frac{\partial}{\partial \rho} \psi(r,\rho), \\ \psi_{\cdot,2}(r,\rho) &= \frac{(N-1)}{8N(N+1)} \frac{\partial^2}{\partial \rho^2} \psi(r,\rho), \end{aligned}\right\} \tag{51}$$

which, being the coefficients of $L_{4,1}$ and $L_{4,2}$ in (32), will be known as corrective factors due to their population values since, when the population is normal, these L-coefficients vanish.

Similarly,

$$\left.\begin{aligned}\psi_{6,1}(r,\rho) &= \frac{(N-2)}{12N(N+1)(N+3)}\frac{\partial}{\partial\rho}\psi(r,\rho),\\ \psi_{6,2}(r,\rho) &= \frac{(N-2)}{12N(N+1)(N+3)}\frac{\partial^2}{\partial\rho^2}\psi(r,\rho),\\ \psi_{6,3}(r,\rho) &= \frac{(N-2)}{12N(N+1)(N+3)}\frac{\partial^3}{\partial\rho^3}\psi(r,\rho),\end{aligned}\right\} \tag{52}$$

will be known as the corrective factors due to $L_{6,1}$, $L_{6,2}$ and $L_{6,3}$ of the population respectively.

The values of the ordinates of the corrective factors (51) and (52) are given in Table 1, for $N = 11$, $\rho = 0{\cdot}0$, and for $N = 11$ and 21, $\rho = 0{\cdot}8$. David's (1938) tables of the correlation coefficient have been used for the purpose of the tabulation work. The data of Table 1 below have been given at some length because they provide a point from which one can start to explore the distribution of r for different populations. The selection of two illustrative values of N and ρ is, as such, useful.

It is seen from the results of Table 1 that in the null case, i.e. when $\rho = 0$, the effect of the parent non-normality is not serious, even for values of N as low as 11. For the case $\rho = 0{\cdot}8$, on the other hand, the effects are not negligible. In particular, the corrective factors $\psi_{4,2}(r,\rho)$ and $\psi_{6,3}(r,\rho)$ seem always to have a considerable part in distorting the normal-theory frequency curve of r, even when the sample size is gradually increased.

However, the magnitude of the effective corrections is also dependent on the population values of the L-coefficients, (33) and (34). Investigations into possible limits of these, for given values of ρ, have not produced any very useful results. It is worth noting that two of the L-coefficients, namely $L_{4,1}$ and $L_{4,2}$, occur in the well-known first approximations to the mean and variance of r, for any population, as corrective terms to their corresponding normal-theory values, thus

$$\begin{aligned}\text{mean } r &\simeq \rho + \frac{\rho}{8N}\left\{3\left(\frac{\mu_{40}}{\mu_{20}^2}+\frac{\mu_{04}}{\mu_{02}^2}\right)-4\left(\frac{\mu_{31}}{\mu_{11}\mu_{20}}+\frac{\mu_{13}}{\mu_{11}\mu_{02}}\right)+2\frac{\mu_{22}}{\mu_{20}\mu_{02}}\right\}\\ &= \rho+\frac{1}{N}\{-\tfrac{1}{2}\rho(1-\rho^2)+\tfrac{1}{8}L_{4,1}\},\end{aligned} \tag{53}$$

and

$$\begin{aligned}\text{var } r &\simeq \frac{\rho^2}{4N}\left\{\frac{\mu_{40}}{\mu_{20}^2}+\frac{\mu_{04}}{\mu_{02}^2}-4\left(\frac{\mu_{31}}{\mu_{11}\mu_{20}}+\frac{\mu_{13}}{\mu_{11}\mu_{02}}\right)+2\left(\frac{\mu_{22}}{\mu_{20}\mu_{02}}+\frac{2\mu_{22}}{\mu_{11}^2}\right)\right\}\\ &= \frac{1}{N}\{(1-\rho^2)^2+\tfrac{1}{4}L_{4,2}\}.\end{aligned} \tag{54}$$

Accordingly, the mean is affected by the population value of $L_{4,1}$ and the variance by that of $L_{4,2}$. Also, since $\lim_{N\to\infty}(N \text{ var } r) \geqslant 0$,

$$L_{4,2} \geqslant -4(1-\rho^2)^2, \tag{55}$$

an inequality which sets a lower limit to $L_{4,2}$ for a given value of the population correlation coefficient ρ.

Table 1. *Corrective functions for the frequency distribution of the correlation coefficient r in samples from non-normal populations*

r	Case: $\rho = 0{\cdot}0$, $N = 11$. The functions give ordinates multiplied by 1000					
	Normal-theory frequency function	Corrective functions for excess		Corrective functions for skewness		
	$\psi(r, 0)$	$\psi_{4,1}(r, 0)$	$\psi_{4,2}(r, 0)$	$\psi_{6,1}(r, 0)$	$\psi_{6,2}(r, 0)$	$\psi_{6,3}(r, 0)$
−1·00	0·00	0·00	0·00	0·00	0·00	0·00
−0·95	0·34	− 0·03	0·28	0·00	0·01	− 0·07
−0·90	3·48	− 0·28	2·32	−0·01	0·10	− 0·84
−0·85	13·10	− 1·00	7·72	−0·04	0·33	− 2·47
−0·80	32·59	− 2·35	16·66	−0·10	0·71	− 4·76
−0·75	64·48	− 4·36	28·24	−0·19	1·21	− 7·10
−0·70	110·28	− 6·95	40·73	−0·30	1·75	− 8·73
−0·65	170·38	− 9·98	52·04	−0·43	2·23	− 9·03
−0·60	244·13	−13·20	60·11	−0·57	2·58	− 7·67
−0·55	329·91	−16·35	63·26	−0·70	2·71	− 4·62
−0·50	425·31	−19·16	60·41	−0·82	2·59	− 0·20
−0·45	527·29	−21·38	51·18	−0·92	2·19	5·04
−0·40	632·37	−22·79	35·93	−0·98	1·54	10·39
−0·35	736·81	−23·23	15·70	−1·00	0·67	15·11
−0·30	836·83	−22·62	− 7·92	−0·97	−0·34	18·52
−0·25	928·73	−20·92	− 32·98	−0·90	−1·41	20·11
−0·20	1009·12	−18·18	− 57·34	−0·78	−2·46	19·60
−0·15	1074·98	−14·53	− 78·89	−0·62	−3·38	16·98
−0·10	1123·87	−10·12	− 95·78	−0·43	−4·11	12·49
−0·05	1153·95	− 5·20	−106·54	−0·22	−4·57	6·62
0	1164·10	0·00	−110·24	0·00	−4·72	0·00
0·05	1153·95	5·20	−106·54	0·22	−4·57	− 6·62
0·10	1123·87	10·12	− 95·78	0·43	−4·11	−12·49
0·15	1074·98	14·53	− 78·89	0·62	−3·38	−16·98
0·20	1009·12	18·18	− 57·34	0·78	−2·46	−19·60
0·25	928·73	20·92	− 32·98	0·90	−1·41	−20·11
0·30	836·83	22·62	− 7·92	0·97	−0·34	−18·52
0·35	736·81	23·23	15·70	1·00	0·67	−15·11
0·40	632·37	22·79	35·93	0·98	1·54	−10·39
0·45	527·29	21·38	51·18	0·92	2·19	− 5·04
0·50	425·31	19·16	60·41	0·82	2·59	0·20
0·55	329·91	16·35	63·26	0·70	2·71	4·62
0·60	244·13	13·20	60·11	0·57	2·58	7·67
0·65	170·38	9·98	52·04	0·43	2·23	9·03
0·70	110·28	6·95	40·73	0·30	1·75	8·73
0·75	64·48	4·36	28·24	0·19	1·21	7·10
0·80	32·59	2·35	16·66	0·10	0·71	4·76
0·85	13·10	1·00	7·72	0·04	0·33	2·47
0·90	3·48	0·28	2·32	0·01	0·10	0·84
0·95	0·34	0·03	0·28	0·00	0·01	0·07
1·00	0·00	0·00	0·00	0·00	0·00	0·00

Table 1 (*continued*)

r	Case: $\rho = 0{\cdot}8$, $N = 11$. The functions give ordinates multiplied by 1000					
	Normal-theory frequency function	Corrective functions for excess		Corrective functions for skewness		
	$\psi(r, \rho)$	$\psi_{4,1}(r, \rho)$	$\psi_{4,2}(r, \rho)$	$\psi_{6,1}(r, \rho)$	$\psi_{6,2}(r, \rho)$	$\psi_{6,3}(r, \rho)$
−0·75	0·00	0·00	0·00	0·00	0·00	0·00
−0·70	0·01	0·00	0·03	0·00	0·00	− 0·02
−0·65	0·02	− 0·01	0·13	0·00	0·01	− 0·09
−0·60	0·04	− 0·01	0·18	0·00	0·01	− 0·11
−0·55	0·06	− 0·01	0·30	0·00	0·01	− 0·21
−0·50	0·10	− 0·02	0·50	0·00	0·02	− 0·34
−0·45	0·17	− 0·04	0·83	0·00	0·04	− 0·55
−0·40	0·27	− 0·06	1·29	0·00	0·06	− 0·82
−0·35	0·43	− 0·10	1·99	0·00	0·09	− 1·24
−0·30	0·65	− 0·15	2·93	− 0·01	0·13	− 1·77
−0·25	0·99	− 0·23	4·31	− 0·01	0·18	− 2·52
−0·20	1·49	− 0·34	6·25	− 0·01	0·27	− 3·50
−0·15	2·21	− 0·49	8·92	− 0·02	0·38	− 4·77
−0·10	3·27	− 0·72	12·62	− 0·03	0·54	− 6·39
−0·05	4·80	− 1·03	17·63	− 0·04	0·76	− 8·39
0·00	7·04	− 1·48	24·49	− 0·06	1·05	− 10·81
0·05	10·29	− 2·12	33·66	− 0·09	1·44	− 13·59
0·10	15·02	− 3·01	45·87	− 0·13	1·97	− 16·61
0·15	21·93	− 4·28	61·91	− 0·18	2·65	− 19·47
0·20	32·04	− 6·06	82·62	− 0·26	3·54	− 21·50
0·25	46·90	− 8·55	108·79	− 0·37	4·66	− 21·45
0·30	68·83	− 12·04	140·78	− 0·52	6·03	− 17·22
0·35	101·35	− 16·89	177·98	− 0·72	7·63	− 5·53
0·40	149·79	− 23·59	217·38	− 1·01	9·32	18·62
0·45	222·30	− 32·71	251·63	− 1·40	10·78	62·01
0·50	331·21	− 44·85	265·07	− 1·92	11·36	132·32
0·55	495·09	− 60·41	228·01	− 2·59	9·77	234·63
0·60	741·14	− 78·97	89·30	− 3·38	3·83	361·19
0·625	906·37	− 88·74	− 41·30	− 3·80	− 1·77	422·36
0·650	1107·09	− 98·01	− 227·73	− 4·20	− 9·76	468·29
0·675	1349·30	− 105·70	− 481·08	− 4·53	− 20·62	481·08
0·700	1638·62	− 110·15	− 808·55	− 4·72	− 34·65	434·53
0·725	1979·08	− 108·95	− 1207·69	− 4·67	− 51·76	294·12
0·750	2370·71	− 98·76	− 1656·60	− 4·23	− 71·00	20·32
0·775	2805·67	− 75·36	− 2098·42	− 3·23	− 89·93	− 420·67
0·800	3261·77	− 33·97	− 2422·26	− 1·46	− 103·81	− 1030·28
0·825	3693·13	29·41	− 2444·64	1·26	− 104·77	− 1730·21
0·850	4017·99	115·07	− 1916·42	4·93	− 82·13	− 2290·31
0·875	4108·41	213·92	− 602·72	9·17	− 25·83	− 2275·78
0·900	3796·61	299·35	1471·74	12·83	63·07	− 1154·60
0·925	2936·70	322·29	3559·76	13·81	152·56	1126·23
0·950	1590·43	231·87	3905·79	9·94	167·39	2975·12
0·975	336·04	63·31	1450·35	2·71	62·16	1618·24
1·000	0·00	0·00	0·00	0·00	0·00	0·00

Table 1 (*continued*)

r	Case: $\rho = 0{\cdot}8$, $N = 21$. The functions give ordinates multiplied by 1000					
	Normal-theory frequency function	Corrective functions for excess		Corrective functions for skewness		
	$\psi(r, \rho)$	$\psi_{4,1}(r, \rho)$	$\psi_{4,2}(r, \rho)$	$\psi_{3,1}(r, \rho)$	$\psi_{3,2}(r, \rho)$	$\psi_{3,3}(r, \rho)$
−0·25	0·00	0·00	0·00	0·00	0·00	0·00
−0·20	0·00	0·00	0·00	0·00	0·00	0·00
−0·15	0·01	0·00	0·09	0·00	0·00	− 0·08
−0·10	0·01	0·00	0·11	0·00	0·00	− 0·10
−0·05	0·03	− 0·01	0·29	0·00	0·01	− 0·26
0·00	0·06	− 0·01	0·56	0·00	0·01	− 0·48
0·05	0·14	− 0·03	1·24	0·00	0·03	− 1·03
0·10	0·29	− 0·07	2·40	0·00	0·06	− 1·90
0·15	0·63	− 0·14	4·92	0·00	0·13	− 3·70
0·20	1·33	− 0·29	9·58	− 0·01	0·25	− 6·72
0·25	2·82	− 0·59	18·62	− 0·02	0·49	− 12·09
0·30	5·95	− 1·18	35·42	− 0·03	0·93	− 20·85
0·35	12·53	− 2·37	66·05	− 0·06	1·74	− 34·33
0·40	26·36	− 4·70	119·95	− 0·12	3·17	− 52·59
0·45	55·31	− 9·20	209·75	− 0·24	5·54	− 71·51
0·50	115·56	− 17·63	347·20	− 0·47	9·16	− 76·26
0·55	239·50	− 32·77	525·23	− 0·86	13·86	− 29·42
0·60	489·12	− 58·07	672·94	− 1·53	17·76	136·32
0·625	692·81	− 75·22	674·81	− 1·99	17·81	287·09
0·650	972·87	− 94·94	567·99	− 2·51	14·99	485·31
0·675	1350·45	− 115·65	287·63	− 3·05	7·59	710·72
0·700	1845·86	− 134·01	− 236·95	− 3·54	− 6·25	908·32
0·725	2471·22	− 144·14	− 1053·55	− 3·80	− 27·80	972·37
0·750	3216·90	− 137·12	− 2132·03	− 3·62	− 56·26	749·64
0·775	4030·04	− 101·83	− 3274·81	− 2·69	− 86·42	81·42
0·800	4787·20	− 28·65	− 4029·94	− 0·76	− 106·35	− 1059·41
0·825	5274·55	82·12	− 3710·40	2·17	− 97·91	− 2341·78
0·850	5209·97	207·44	− 1729·79	5·47	− 45·65	− 2914·05
0·875	4364·79	293·73	1593·30	7·75	42·05	− 1781·58
0·900	2814·91	277·84	4281·50	7·33	112·98	821·21
0·925	1157·18	156·15	3854·73	4·12	101·72	2362·23
0·950	195·96	34·73	1192·94	0·92	31·48	1124·91
0·975	3·29	0·75	33·32	0·02	0·88	44·26
1·000	0·00	0·00	0·00	0·00	0·00	0·00

The fact that the expression on the right-hand side of (54) is always positive also follows directly from the following consideration. By writing $\mu_{20} = \sigma_1^2, \mu_{02} = \sigma_2^2$ and $\mu_{11} = \rho\sigma_1\sigma_2$, it will be seen that

$$\operatorname{var} r \simeq \frac{1}{4N} E\left[\rho^2\left\{\left(\frac{x-\xi}{\sigma_1}\right)^4 + \left(\frac{y-\eta}{\sigma_2}\right)^4\right\} - 4\rho\left\{\left(\frac{x-\xi}{\sigma_1}\right)^3\left(\frac{y-\eta}{\sigma_2}\right) + \left(\frac{x-\xi}{\sigma_1}\right)\left(\frac{y-\eta}{\sigma_2}\right)^3\right\} + 2(2+\rho^2)\left(\frac{x-\xi}{\sigma_1}\right)^2\left(\frac{y-\eta}{\sigma_2}\right)^2\right]$$

$$= \frac{1}{4N} E\left\{\rho\left(\frac{x-\xi}{\sigma_1}\right)^2 - 2\left(\frac{x-\xi}{\sigma_1}\right)\left(\frac{y-\eta}{\sigma_2}\right) + \rho\left(\frac{y-\eta}{\sigma_2}\right)^2\right\}^2$$

$$\geqslant 0, \qquad (54)\,bis$$

where ξ and η are the population means of x and y.

Obviously, neither the L-coefficients nor the semi-invariants themselves can take arbitrary values. For various inequality relations are known to exist between similar order semi-invariants of a population. In order to see how large these L-coefficients can be in practice, we shall consider a few numerical illustrations from the studies of various workers. With a short description of the source of data, Table 2 shows the semi-invariants of a few non-normal distributions, arranged in order of magnitude of the correlation coefficient ρ. As the number of observations in all these examples is large, it is not unlikely that the calculated semi-invariants are truly representative of their population values. In some of the cases the Edgeworth surface (1) may not afford a very satisfactory representation of the population. But whatever may be the population form, as the effect of the first power of the fourth-order and the square and product of the third-order semi-invariants can be correctly measured by the factors (51) and (52), it seems worth while to examine the values of the L-coefficients for all these populations. They are shown in the last columns of Table 2.

The values of the semi-invariants for populations I (no. of trumps in deals of whist) and III (weights of placenta and newborn child (boys)) as given by Wicksell, are correct up to four figures only. So that in the present, as well as in the subsequent illustrations, the calculated values of the constants for these populations may be in error in the last one or two figures.

The population of contemporaneous barometric heights at Southampton and Laudale (no. IV in Table 2) might be considered as moderately non-normal. Pearson (1925) fitted his fifteen constant bivariate frequency surface (which is the same as the bivariate Gram-Charlier surface including terms as far as those in $\lambda_{40}, \lambda_{31}, \lambda_{22}, \lambda_{13}$ and λ_{04} only) to the correlation table of this distribution. Regarding the goodness of fit, he says, 'If the 15-constant surface actually describes the material sampled, we should expect a worse result than that observed in one out of five or six trials. Such a result is sufficiently encouraging to make it worthwhile investigating the 15-constant surface on further material.' The Edgeworth population (1), on the other hand, as it considers a further set of terms in $\lambda_{30}^2, \lambda_{30}\lambda_{21} \ldots, \lambda_{03}\lambda_{12}$ and λ_{03}^2, will certainly have higher representational power in such cases than the 15-constant surface. In what follows, we shall investigate in some detail, by applying our theoretical formulae, the sampling distribution of r for such a parent population. The measures of departure from normality being moderate, our formulae should work well in this case.

Table 2. *The population parameters of a few non-normal distributions and the corresponding values of the L-coefficients*

	Description, no. of observations and source of data	Population value of ρ	Population semi-invariants									Corrective coefficients from equation (33)		Corrective coefficients from equation (34) with $N = 11$		
			Third order				Fourth order									
			λ_{30}	λ_{21}	λ_{12}	λ_{03}	λ_{40}	λ_{31}	λ_{22}	λ_{13}	λ_{04}	$L_{4,1}$	$L_{4,2}$	$L_{5,1}$	$L_{5,2}$	$L_{5,3}$
I	No. of trumps in deals of whist in two hands—25,000 (Pearson-Isserlis, 1914)	−0·3305	0·1826	−0·0614	−0·0627	0·1857	−0·0653	0·0244	0·0266	0·0079	−0·0466	−0·2577	0·1428	−0·0212	0·0357	0·1160
II	Barometric heights on alternate days—28,855 (full year) (Pretorius, 1930)	0·5807	0·4508	0·1521	0·1694	0·4513	0·3978	0·1392	0·1426	0·1232	0·3988	0·5036	0·3255	−0·9064	−1·0200	0·2178
III	Weights of newborn child and placenta (boys)—1223 (Wicksell, 1917)	0·6455	−0·1320	0·0641	0·1506	0·4824	0·2856	−0·0080	−0·0683	−0·2068	−0·1106	1·1100	0·2974	−1·2778	−1·1174	0·1207
IV	Contemporaneous barometric heights —2922 Southampton and Laudale (Pearson, 1925)	0·7808	0·4137	0·2870	0·3104	0·4739	0·6120	0·3341	0·2250	0·1904	0·1949	0·1437	0·0282	0·0256	−0·4995	0·0415
V	Length and breadth of beans—9440 (Wicksell, 1917, after Johansen)	0·7811	−0·9106	−0·6533	−0·5025	−0·4408	1·8629	1·2984	0·9466	0·7292	0·6544	−0·7323	0·1425	1·9209	−0·6409	0·0335

Illustration

For the frequency distribution IV of Table 2, Karl Pearson has given the following values of q_{ij} (see expression (40) above), for $i+j = 3$ and 4:

$$\rho = q_{11} = 0{\cdot}780{,}819$$

$$q_{30} = 0{\cdot}413{,}691 \qquad q_{40} = 3{\cdot}612{,}028$$
$$q_{21} = 0{\cdot}286{,}962 \qquad q_{31} = 2{\cdot}676{,}586$$
$$q_{12} = 0{\cdot}310{,}386 \qquad q_{22} = 2{\cdot}444{,}352$$
$$q_{03} = 0{\cdot}473{,}852 \qquad q_{13} = 2{\cdot}532{,}831$$
$$q_{04} = 3{\cdot}194{,}947.$$

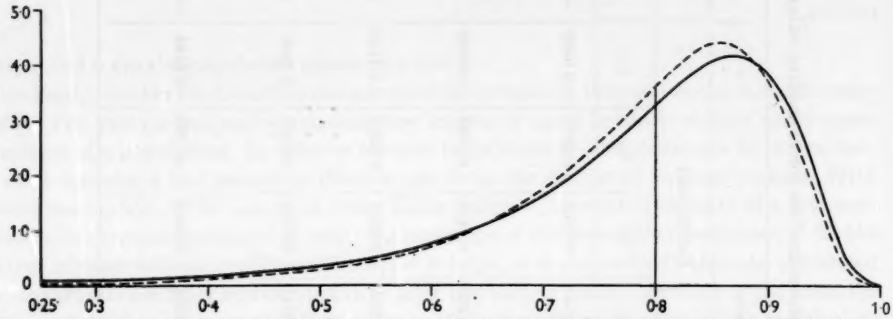


Fig. 1*a*. The frequency curve of r ($N = 11$, $\rho = 0{\cdot}8$) in samples from: (i) a normal population, —; (ii) a non-normal population (based on contemporaneous barometric heights distribution), – – –.

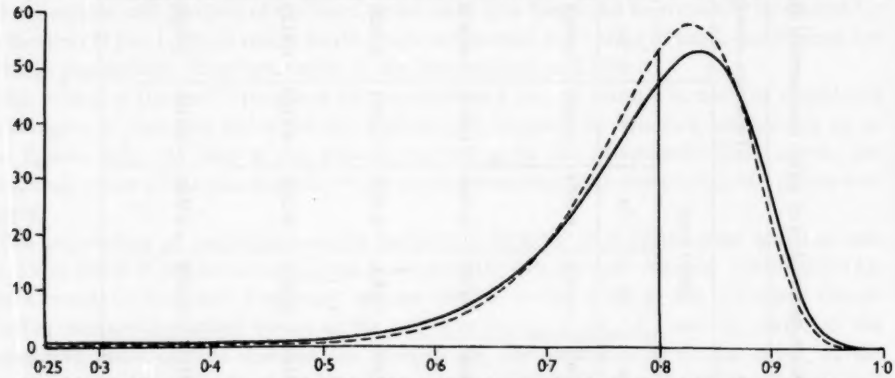


Fig. 1*b*. The frequency curve of r ($N = 21$, $\rho = 0{\cdot}8$) in samples from: (i) a normal population, —; (ii) a non-normal population (based on contemporaneous barometric heights distribution), – – –.

To cut down the numerical work, we shall take

$$\rho = q_{11} = 0{\cdot}8$$

$$\lambda_{30} = q_{30} = 0{\cdot}4 \qquad \lambda_{40} = q_{40} - 3 = 0{\cdot}6$$
$$\lambda_{21} = q_{21} = 0{\cdot}3 \qquad \lambda_{31} = q_{31} - 3\rho = 0{\cdot}3$$
$$\lambda_{12} = q_{12} = 0{\cdot}3 \qquad \lambda_{22} = q_{22} - 2\rho^2 - 1 = 0{\cdot}12$$
$$\lambda_{03} = q_{03} = 0{\cdot}5 \qquad \lambda_{13} = q_{13} - 3\rho = 0{\cdot}1$$
$$\lambda_{04} = q_{04} - 3 = 0{\cdot}2$$

and proceed to investigate the theoretical distribution of r, in samples of $N = 11$ and $N = 21$ for the population specified by the above set of values of ρ and λ's. The approximations thus considered will enable us to utilize our tabulated values of the ordinates of the corrective functions for $\rho = 0{\cdot}8$ and $N = 11$ and 21.

Table 3. *The probability integrals of the frequency distribution of r, derived from formula* (32), *in samples from the non-normal population IV, based on the frequency of contemporaneous barometric heights with the corresponding normal-theory values*

r	Probability integral for the distribution of r, ($\rho = 0{\cdot}8$)			
	$N = 11$		$N = 21$	
	Population IV	Normal population	Population IV	Normal population
−0·10	0·00006	0·00040	0·00000	0·00000
0·00	0·00019	0·00089	0·00000	0·00000
0·10	0·00056	0·00194	0·00000	0·00002
0·20	0·00154	0·00419	0·00000	0·00009
0·30	0·00413	0·00900	0·00000	0·00040
0·40	0·01085	0·01940	0·00011	0·00177
0·50	0·02830	0·04223	0·00215	0·00781
0·60	0·07364	0·09311	0·02052	0·03383
0·65	0·11807	0·13872	0·04945	0·06911
0·70	0·18787	0·20658	0·11416	0·13761
0·75	0·29451	0·30596	0·24439	0·26217
0·80	0·44876	0·44643	0·46649	0·46323
0·825	0·54432	0·53348	0·60731	0·58983
0·850	0·64901	0·63020	0·75058	0·72237
0·875	0·75666	0·73243	0·87414	0·84377
0·900	0·85650	0·83226	0·95766	0·93457
0·925	0·93573	0·91764	0·99449	0·98353
0·950	0·98254	0·97489	0·99914	0·99848
0·975	0·99876	0·99776	0·99999	0·99998
1·000	1·00000	1·00000	1·00000	1·00000

In order to avoid confusion in the latter part of this work, we shall denote Karl Pearson's (1925) distribution of contemporaneous barometric heights as the 'Population IV_0' and the present approximated distribution based on the actual one as the 'Population IV'.

Now the L-coefficients for this population IV are found to be

$$L_{4,1} = 0{\cdot}5120, \quad L_{4,2} = -0{\cdot}1344$$

$$L_{6,1} = -0{\cdot}2053{,}333 \text{ for } N = 11, \text{ and } -0{\cdot}2425{,}263 \text{ for } N = 21,$$

$$L_{6,2} = -0{\cdot}5216{,}000 \text{ for } N = 11, \text{ and } \quad 0{\cdot}4668{,}631 \text{ for } N = 21,$$

$$L_{6,3} = \quad 0{\cdot}0307{,}200 \text{ for } N = 11, \text{ and } \cdot 0{\cdot}0259{,}199 \text{ for } N = 21.$$

Comparative values of the probability integral of the distribution of r, for the normal population and the non-normal population IV will be seen in Table 3. As the tables for the

ordinates and the probability integrals of normal-theory r are available, correct to five decimal figures only, the last one or two digits of the tabulated results in the non-normal case may be in error. Comparative frequency curves of r for the normal and non-normal cases will be seen in Figs. 1*a*, *b*. The population value $\rho = 0{\cdot}8$ has been marked by an ordinate in each diagram.

5. The moment constants and the efficacy of the distribution of $z = \frac{1}{2}\log_e(1+r)/(1-r)$, in samples from a normal population—a correction to some formulae of R. A. Fisher

Suggesting the transformation

$$z = \tfrac{1}{2}\log_e\frac{1+r}{1-r} \quad \text{and} \quad \zeta = \tfrac{1}{2}\log_e\frac{1+\rho}{1-\rho} \tag{56}$$

for the sample correlation coefficient r and its population value ρ, Fisher (1921) gave the asymptotic expansion* for the frequency function of r, in samples from a normal population, in powers of $x = z-\zeta$ and inverse powers of $(N-1)$, N being the size of sample. The expressions for the moment-constants of x about $x = 0$ and $x = \bar{x}$, as well as the first two β-coefficients were also given in that connexion. Unfortunately, an error has led to most of these formulae being incorrect in certain terms. The correct expressions are given below. Thus we have for the raw moments,

$$\mu'_1(z) = \zeta + \frac{\rho}{2(N-1)}\left\{1+\frac{5+\rho^2}{4(N-1)}+\ldots\right\}, \tag{57}$$

$$\mu'_2(x) = \frac{1}{(N-1)}\left\{1+\frac{8-\rho^2}{4(N-1)}+\frac{88-9\rho^2-9\rho^4}{24(N-1)^2}+\ldots\right\}, \tag{58}$$

$$\mu'_3(x) = \frac{3\rho}{2(N-1)^2}\left\{1+\frac{13+2\rho^2}{4(N-1)}+\ldots\right\}, \tag{59}$$

$$\mu'_4(x) = \frac{1}{(N-1)^2}\left\{3+\frac{28-3\rho^2}{2(N-1)}+\frac{736-84\rho^2-51\rho^4}{16(N-1)^2}+\ldots\right\}, \tag{60}$$

for the central moments

$$\mu_2(z) = \frac{1}{(N-1)}\left\{1+\frac{4-\rho^2}{2(N-1)}+\frac{22-6\rho^2-3\rho^4}{6(N-1)^2}+\ldots\right\}, \tag{61}$$

$$\mu_3(z) = \frac{\rho^3}{(N-1)^3}+\ldots, \tag{62}$$

$$\mu_4(z) = \frac{1}{(N-1)^2}\left\{3+\frac{14-3\rho^2}{(N-1)}+\frac{184-48\rho^2-21\rho^4}{4(N-1)^2}+\ldots\right\}, \tag{63}$$

and for the β-coefficients

$$\beta_1(z) = \frac{\rho^6}{(N-1)^3}+\ldots, \tag{64}$$

$$\beta_2(z) = 3+\frac{2}{(N-1)}+\frac{4+2\rho^2-3\rho^4}{(N-1)^2}+\ldots. \tag{65}$$

The formulae in their incorrect form have passed into general use, as most of the workers on the present problem have used them for comparison of their experimental results and the corresponding normal-theory approximation by z-transformation. The corrections affect, for example, Table V of E. S. Pearson's (1929) paper and Table VIII in the Introduction to

* There is an obvious printer's error in the third term of the expression within curly brackets (p. 13). It should be $\frac{2+\rho^2}{8(n-1)}$ instead of $\frac{2+\rho}{8(n-1)}$ (in Fisher's notation).

David's (1938) *Tables of the Distribution of the Correlation Coefficient.* In general, the approximations based on the corrected formulae show that Fisher's z-transformation provides even more accurate results than were suggested by the comparisons quoted by these authors.*

6. The distribution of Fisher's z, in samples from a non-normal population

Let us apply the transformation (56) to the frequency function (32). Writing

$$\chi(z,\zeta)=\frac{(N-2)}{\pi}\operatorname{sech}^{N-1}\zeta\operatorname{sech}^{N-2}z\int_0^\infty\frac{dt}{(\cosh t-\rho r)^{N-1}} \tag{66}$$

which is the normal-theory frequency density of z, the transformed frequency function can be put in the form

$$f(z)=\chi(z,\zeta)+\frac{N-1}{8N(N+1)}\left\{L_{4,1}\frac{\partial}{\partial\rho}\chi(z,\zeta)+L_{4,2}\frac{\partial^2}{\partial\rho^2}\chi(z,\zeta)\right\}$$
$$+\frac{N-2}{12N(N+1)(N+3)}\left\{L_{6,1}\frac{\partial}{\partial\rho}\chi(z,\zeta)+L_{6,2}\frac{\partial^2}{\partial\rho^2}\chi(z,\zeta)+L_{6,3}\frac{\partial^3}{\partial\rho^3}\chi(z,\zeta)\right\}. \tag{67}$$

For the direct evaluation of the moments of z about ζ, from (67), we shall come across the integrals

$$Q(k,\nu)=\int_{-\infty}^{\infty}(z-\zeta)^k\left(\frac{\partial}{\partial\rho}\right)^\nu\{\chi(z,\zeta)\}\,dz, \tag{68}$$

which can be expressed in terms of the derivatives with respect to ρ of the normal-theory moments

$$\mu'_k(x)=\int_{-\infty}^{\infty}(z-\zeta)^k\chi(z,\zeta)\,dz. \tag{69}$$

In the present case, for $\nu=1,2,$ and 3, the integrals (68) will be found to be

$$Q(k,1)=\frac{\partial}{\partial\rho}\mu'_k(x)+\frac{k}{(1-\rho^2)}\mu'_{k-1}(x), \tag{70}$$

$$Q(k,2)=\frac{\partial^2}{\partial\rho^2}\mu'_k(x)+\frac{2k}{(1-\rho^2)}\frac{\partial}{\partial\rho}\mu'_{k-1}(x)+\frac{2k\rho}{(1-\rho^2)^2}\mu'_{k-1}(x)+\frac{k(k-1)}{(1-\rho^2)^2}\mu'_{k-2}(x) \tag{71}$$

and

$$Q(k,3)=\frac{\partial^3}{\partial\rho^3}\mu'_k(x)+\frac{3k}{(1-\rho^2)}\frac{\partial^2}{\partial\rho^2}\mu'_{k-1}(x)+\frac{6\rho k}{(1-\rho^2)^2}\frac{\partial}{\partial\rho}\mu'_{k-1}(x)+\frac{3k(k-1)}{(1-\rho^2)^2}\frac{\partial}{\partial\rho}\mu'_{k-2}(x)$$
$$+\frac{2(1+3\rho^2)k}{(1-\rho^2)^3}\mu'_{k-1}(x)+\frac{6\rho k(k-1)}{(1-\rho^2)^3}\mu'_{k-2}(x)+\frac{k(k-1)(k-2)}{(1-\rho^2)^3}\mu'_{k-3}(x). \tag{72}$$

The asymptotic formulae for the first four moments of z about ζ for the distribution (67) may be obtained by using the corresponding normal-theory results, (57)–(60), in the above integrals (70)–(72). The expressions for the mean, the variance and the third and the fourth moments about the mean will be found to be

$$\text{mean } z=\zeta+\frac{1}{(N-1)}\left\{\tfrac{1}{4}\rho+\frac{1}{8(1-\rho^2)^2}[\rho(3-\rho^2)(\lambda_{40}+\lambda_{04})-4(1+\rho^2)(\lambda_{31}+\lambda_{13})+2\rho(5+\rho^2)\lambda_{22}]\right\}$$
$$+\frac{1}{(N-1)^2}\left\{\tfrac{1}{8}\rho(5+\rho^2)-\frac{1}{16(1-\rho^2)^2}[3\rho(5-\rho^4)(\lambda_{40}+\lambda_{04})-4(5+8\rho^2-\rho^4)(\lambda_{31}+\lambda_{13})\right.$$
$$+2\rho(29+8\rho^2-\rho^4)\lambda_{22}]-\frac{1}{12(1-\rho^2)^3}[\rho(15-10\rho^2+3\rho^4)(\lambda_{30}^2+\lambda_{03}^2)$$
$$+3\rho(15+10\rho^2-\rho^4)(\lambda_{21}^2+\lambda_{12}^2)-4(1+3\rho^2)\lambda_{30}\lambda_{03}-12(2+9\rho^2+\rho^4)\lambda_{21}\lambda_{12}$$
$$\left.-6(3+6\rho^2-\rho^4)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+12\rho(3+\rho^2)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\right\}, \tag{73}$$

* (The corrections involved to Table VIII of David's (1938) *Tables of the Distribution of the Correlation Coefficient* have been kindly supplied by Dr Gayen and will be included on an Errata Sheet issued with copies of these Tables sold in future. ED.)

$$\text{var } z = \frac{1}{(N-1)}\left\{1+\frac{1}{4(1-\rho^2)^2}[\rho^2(\lambda_{40}+\lambda_{04})-4\rho(\lambda_{31}+\lambda_{13})+2(2+\rho^2)\lambda_{22}]\right\}$$
$$+\frac{1}{(N-1)^2}\left\{\tfrac{1}{2}(4-\rho^2)-\frac{1}{4(1-\rho^2)^2}[\rho^2(2+\rho^2)(\lambda_{40}+\lambda_{04})-4\rho(2+\rho^2)(\lambda_{31}+\lambda_{13})\right.$$
$$+2(4+4\rho^2+\rho^4)\lambda_{22}]-\frac{1}{2(1-\rho^2)^3}[\rho^2(3-\rho^2)(\lambda_{30}^2+\lambda_{03}^2)$$
$$+(4+13\rho^2+\rho^4)(\lambda_{21}^2+\lambda_{12}^2)-4\rho\lambda_{30}\lambda_{03}-12\rho(2+\rho^2)\lambda_{21}\lambda_{12}$$
$$\left.-2\rho(5+\rho^2)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+4(1+2\rho^2)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\right\}$$
$$+\frac{1}{(N-1)^3}\left\{\tfrac{1}{6}(22-6\rho^2-3\rho^4)+\frac{1}{8(1-\rho^2)^2}[\rho^2(7+12\rho^2-5\rho^4)(\lambda_{40}+\lambda_{04})\right.$$
$$-4\rho(9+8\rho^2-3\rho^4)(\lambda_{31}+\lambda_{13})+2(20+21\rho^2+4\rho^4-3\rho^6)\lambda_{22}]$$
$$+\frac{1}{2(1-\rho^2)^3}[\rho^2(21-3\rho^2-2\rho^4)(\lambda_{30}^2+\lambda_{03}^2)+(32+99\rho^2+15\rho^4-2\rho^6)(\lambda_{21}^2+\lambda_{12}^2)$$
$$-8\rho(3+\rho^2)\lambda_{30}\lambda_{03}-36\rho(5+3\rho^2)\lambda_{21}\lambda_{12}$$
$$\left.-2\rho(35+15\rho^2-2\rho^4)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+4(6+17\rho^2+\rho^4)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\right\}, \quad (74)$$

$$\mu_3(z) = \frac{1}{(N-1)^2}\left\{\frac{1}{2(1-\rho^2)^3}[-\rho^3(\lambda_{30}^2+\lambda_{03}^2)-3\rho(2+\rho^2)(\lambda_{21}^2+\lambda_{12}^2)+2\lambda_{30}\lambda_{03}\right.$$
$$\left.+6(1+2\rho^2)\lambda_{21}\lambda_{12}+6\rho^2(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})-6\rho(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\right\}$$
$$+\frac{1}{(N-1)^3}\left\{\rho^3-\frac{1}{4(1-\rho^2)}[3\rho^3(\lambda_{40}+\lambda_{04})-12\rho^2(\lambda_{31}+\lambda_{13})+6\rho(2+\rho^2)\lambda_{22}]\right.$$
$$+\frac{1}{4(1-\rho^2)^3}[\rho^3(13+3\rho^2)(\lambda_{30}^2+\lambda_{03}^2)+9\rho(10+5\rho^2+\rho^4)(\lambda_{21}^2+\lambda_{12}^2)$$
$$-2(7+9\rho^2)\lambda_{30}\lambda_{03}-18(5+9\rho^2+2\rho^4)\lambda_{21}\lambda_{12}$$
$$\left.-6\rho^2(13+3\rho^2)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+6\rho(11+5\rho^2)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\right\}, \quad (75)$$

$$\mu_4(z) = \frac{1}{(N-1)^2}\left\{3+\frac{1}{2(1-\rho^2)^2}[3\rho^2(\lambda_{40}+\lambda_{04})-12\rho(\lambda_{31}+\lambda_{13})+6(2+\rho^2)\lambda_{22}]\right\}$$
$$+\frac{1}{(N-1)^3}\left\{14-3\rho^2-\frac{1}{4(1-\rho^2)^2}[9\rho^4(\lambda_{40}+\lambda_{04})-36\rho^3(\lambda_{31}+\lambda_{13})+18\rho^2(2+\rho^2)\lambda_{22}]\right.$$
$$+\frac{1}{(1-\rho^2)^3}[\rho^2(9-\rho^2)(\lambda_{30}^2+\lambda_{03}^2)+3(4+17\rho^2+3\rho^4)(\lambda_{21}^2+\lambda_{12}^2)-16\rho\lambda_{30}\lambda_{03}$$
$$\left.-12\rho(7+5\rho^2)\lambda_{21}\lambda_{12}-6\rho(5+3\rho^2)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+12(1+3\rho^2)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\right\}$$
$$+\frac{1}{(N-1)^4}\left\{\tfrac{1}{4}(184-48\rho^2-21\rho^4)+\frac{1}{4(1-\rho^2)^2}[\rho^2(19+36\rho^2-27\rho^4)(\lambda_{40}+\lambda_{04})\right.$$
$$-4\rho(25+24\rho^2-21\rho^4)(\lambda_{31}+\lambda_{13})+2(56+61\rho^2-12\rho^4-21\rho^6)\lambda_{22}]$$
$$+\frac{1}{2(1-\rho^2)^3}[3\rho^2(30+3\rho^2-7\rho^4)(\lambda_{30}^2+\lambda_{03}^2)+9(16+54\rho^2+11\rho^4-3\rho^6)(\lambda_{21}^2+\lambda_{12}^2)$$
$$-12\rho(9+4\rho^2)\lambda_{30}\lambda_{03}-36\rho(23+17\rho^2-\rho^4)\lambda_{21}\lambda_{12}$$
$$\left.-6\rho(50+37\rho^2-9\rho^4)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+12(8+30\rho^2+\rho^4)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\right\}, \quad (76)$$

from which we obtain, omitting terms of negligible order,

$$\beta_1(z) = \frac{\rho^3}{(N-1)^2}\Big\{\frac{1}{(1-\rho^2)^3}[-\rho^3(\lambda_{30}^2+\lambda_{03}^2)-3\rho(2+\rho^2)(\lambda_{21}^2+\lambda_{12}^2)+2\lambda_{30}\lambda_{03}$$
$$+6(1+2\rho^2)\lambda_{21}\lambda_{12}+6\rho^2(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})-6\rho(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]$$
$$+\frac{1}{(N-1)}\Big(\rho^3-\frac{3\rho(2-\rho^2)}{4(1-\rho^2)^2}[\rho^2(\lambda_{40}+\lambda_{04})-4\rho(\lambda_{31}+\lambda_{13})+2(2+\rho^2)\lambda_{22}]$$
$$+\frac{1}{2(1-\rho^2)^3}[25\rho^3(\lambda_{30}^2+\lambda_{03}^2)+9\rho(18+7\rho^2)(\lambda_{21}^2+\lambda_{12}^2)$$
$$-2(19+6\rho^2)\lambda_{30}\lambda_{03}-18(9+16\rho^2)\lambda_{21}\lambda_{12}$$
$$-150\rho^2(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+6\rho(23+2\rho^2)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\Big)\Big\}, \tag{77}$$

and

$$\beta_2(z) = 3+\frac{1}{(N-1)}\Big\{2-\frac{1}{(1-\rho^2)^2}[\rho^2(\lambda_{40}+\lambda_{04})-4\rho(\lambda_{31}+\lambda_{13})+2(2+\rho^2)\lambda_{22}]$$
$$+\frac{1}{(1-\rho^2)^3}[2\rho^2(9-2\rho^2)(\lambda_{30}^2+\lambda_{03}^2)+6(4+15\rho^2+2\rho^4)(\lambda_{21}^2+\lambda_{12}^2)$$
$$-28\rho\lambda_{30}\lambda_{03}-12\rho(13+8\rho^2)\lambda_{21}\lambda_{12}$$
$$-12\rho(5+2\rho^2)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})+12(2+5\rho^2)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\Big\}$$
$$+\frac{1}{(N-1)^2}\Big\{4+2\rho^2-3\rho^4+\frac{(4+5\rho^2-3\rho^4)}{2(1-\rho^2)^2}[\rho^2(\lambda_{40}+\lambda_{04})-4\rho(\lambda_{31}+\lambda_{13})+2(2+\rho^2)\lambda_{22}]$$
$$-\frac{1}{(1-\rho^2)^3}[\rho^2(66-35\rho^2+7\rho^4)(\lambda_{30}^2+\lambda_{03}^2)$$
$$+(88+292\rho^2-35\rho^4-3\rho^6)(\lambda_{21}^2+\lambda_{12}^2)-2\rho(49-11\rho^2)\lambda_{30}\lambda_{03}$$
$$-6\rho(103+17\rho^2-6\rho^4)\lambda_{21}\lambda_{12}-2\rho(110+7\rho^2-3\rho^4)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})$$
$$+2(44+91\rho^2-21\rho^4)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21})]\Big\}. \tag{78}$$

The corrective terms in the expressions for the moment constants of z, as they contain powers of $(1-\rho^2)$ in the denominator, may appear at first sight not to be valid near $\rho = 1$. But the values taken by the expressions in the square brackets in the numerators will prevent the whole terms from becoming large. For it will be seen that for $\rho = 1$, the numerators of all corrective terms reduce to one or other of the forms

$$\text{const.}\,(\lambda_{40}-4\lambda_{31}+6\lambda_{22}-4\lambda_{13}+\lambda_{04}),$$

and

$$\text{const.}\,(\lambda_{30}-3\lambda_{21}+3\lambda_{12}-\lambda_{03})^2,$$

which also are known (Kendall, 1949, result (26)) to be zero, in the case $\rho = 1$. It may be possible, however, to find limits of the whole terms for given values of ρ, but investigations on these lines have not been fruitful. It will be noticed that the above corrective terms bear interesting analogies to those in the expressions for the mean and variance of t (the sample value of Kendall's τ) given by Kendall (1949).

7. The effect of the population excess and skewness on the distribution of z

As before, we shall write

$$\left.\begin{aligned}\chi_{4,1}(z,\zeta) &= \frac{(N-1)}{8N(N+1)}\frac{\partial}{\partial\rho}\chi(z,\zeta),\\ \chi_{4,2}(z,\zeta) &= \frac{(N-1)}{8N(N+1)}\frac{\partial^2}{\partial\rho^2}\chi(z,\zeta),\end{aligned}\right\} \tag{79}$$

for the corrective factors in the distribution of z, (67), due to the population values of the constants $L_{4,1}$ and $L_{4,2}$; and

$$\left.\begin{aligned}\chi_{6,1}(z,\zeta) &= \frac{(N-2)}{12N(N+1)(N+3)}\frac{\partial}{\partial\rho}\chi(z,\zeta),\\ \chi_{6,2}(z,\zeta) &= \frac{(N-2)}{12N(N+1)(N+3)}\frac{\partial^2}{\partial\rho^2}\chi(z,\zeta),\\ \chi_{6,3}(z,\zeta) &= \frac{(N-2)}{12N(N+1)(N+3)}\frac{\partial^3}{\partial\rho^3}\chi(z,\zeta),\end{aligned}\right\} \tag{80}$$

for those due to the values of $L_{6,1}$, $L_{6,2}$ and $L_{6,3}$, which, as has been pointed out, depend on the sample size as well.

The total effect of the population excess and skewness on the distribution of z will be studied in some detail by the help of the moment constants deduced in (73)–(78). The expressions for $\beta_1(z)$ and $\beta_2(z)$ show that the distribution of z is asymptotically normal, although the approach towards normality is not so rapid as in the normal theory case. It will be noticed that the expression for $\beta_1(z)$, in (77), begins with a term in $(N-1)^{-2}$, whereas in the normal-theory case it starts with the next higher order term, namely that in $(N-1)^{-3}$. This effect on $\beta_1(z)$ is due to skewness in the population, as measured by $\lambda_{30}^2, \lambda_{30}\lambda_{21}, \ldots, \lambda_{03}\lambda_{12}$ and λ_{03}^2. In both the normal and non-normal cases, however, $\beta_1(z)$ vanishes absolutely for $\rho = 0$. Considering terms up to $O[(N-1)^{-1}]$, the value of $\beta_2(z)$, (78), is found to be influenced by the correlation and also the excess and skewness of the population; whereas in the normal-theory case it does not even depend on ρ.

These observations, it may be emphasized at this point, apply only for moderately non-normal populations. Lower order terms in $(N-1)^{-1}$, containing powers of $\lambda_{30}^2, \lambda_{30}\lambda_{21}, \ldots, \lambda_{03}^2$; $\lambda_{40}, \lambda_{31}, \ldots, \lambda_{04}$, have not been taken into account in the above expressions for the β's. If in particular the squares and products of $\lambda_{30}^2, \lambda_{30}\lambda_{21}, \ldots, \lambda_{03}^2$ of a population are not small, the value of $\beta_1(z)$ may be considerable as in that case its expression will begin with a term in $(N-1)^{-1}$. Accordingly, however large the sample size is, the formula (77) holds good only in cases where population λ's other than those considered are negligibly small.

For the values of ρ given for the five populations listed in Table 2, and for $N \geqslant 21$, $\beta_1(z)$ is always $\leqslant 0{\cdot}00003$ in samples from normal populations. Using equations (74) and (75) with the λ's of Table 2, it was found for $N \geqslant 21$ that while $\beta_1(z)$ had larger values than on the 'normal-theory' it was still always below 0·002. For $\gamma_2(z) = \beta_2(z) - 3$, the comparison is summarized in Table 4.

As the illustrations taken are representative of a fairly wide class of non-normal populations, it may be inferred from the tabulated results that the distribution of z is still in general closely approximated by the normal curve if the samples are not too small.

Considering terms up to $O[(N-1)^{-1}]$, the mean and variance of z, (73) and (74), are independent of the population skewness, showing that its effect on them is less serious. They are mainly influenced by the excess of the population. The mean value of z exceeds the population value by about

$$\frac{1}{(N-1)}\left\{\tfrac{1}{2}\rho + \frac{1}{8(1-\rho^2)^2}[\rho(3-\rho^2)(\lambda_{40}+\lambda_{04}) - 4(1+\rho^2)(\lambda_{31}+\lambda_{13}) + 2\rho(5+\rho^2)\lambda_{22}]\right\}.$$

Compared to $\frac{1}{2}\rho$, the values of the second term of the expression within the curly bracket are not small for the populations of Table 2, showing that the bias introduced by

Table 4. *The comparative values of* $\gamma_2(z) = \beta_2(z) - 3$ *of the distribution of* z, *in samples from normal and non-normal populations*

	Population	Population value of ρ	Values of $\gamma_2(z) = \beta_2(z) - 3$ from formula (78)				
			$N = 21$	$N = 51$	$N = 101$	$N = 201$	$N = 401$
I	Trumps	−0·3305	0·128	0·050	0·025	0·012	0·006
II	Full year barometric heights	0·5807	0·043	0·011	0·004	0·002	0·001
III	Newborn child	0·6455	0·282	0·118	0·060	0·030	0·015
IV_0	Contemporaneous barometric heights*	0·7808	0·130	0·036	0·015	0·007	0·003
V	Beans	0·7811	0·362	0·099	0·042	0·019	0·009
Normal population with the values of ρ as above			0·111 or 0·110	0·042	0·020	0·010	0·005

* Note that the figures in this row are for the original population IV_0 (see p. 235 above).

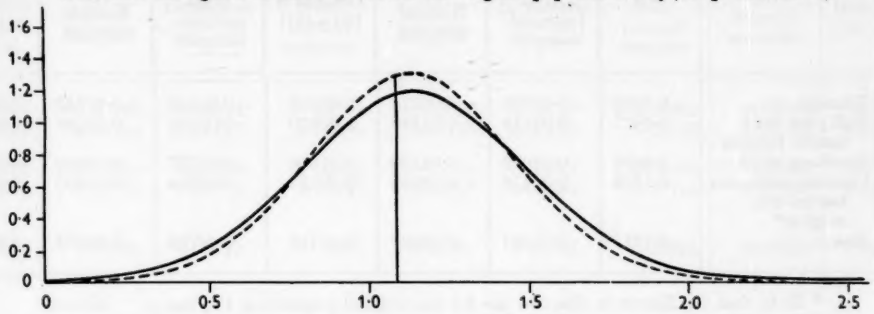


Fig. 2*a*. Comparative frequency curves for $z = \frac{1}{2}\log_e \frac{1+r}{1-r}$, $(N = 11, \rho = 0{\cdot}8)$ in samples from: (i) a normal population, —; (ii) a non-normal population (based on contemporaneous barometric heights distribution), – – –.

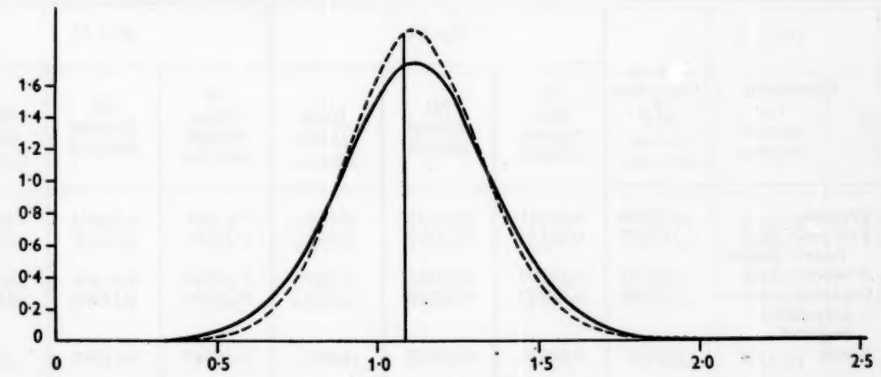


Fig. 2*b*. Comparative frequency curves for $z = \frac{1}{2}\log_e \frac{1+r}{1-r}$, $(N = 21, \rho = 0{\cdot}8)$ in samples from: (i) a normal population, —; (ii) a non-normal population (based on contemporaneous barometric heights distribution), – – –.

non-normality may be considerable. Using the full expressions (73) and (74) the values of mean $z-\zeta$, and standard deviation of z, (σ_z), have been calculated for a few representative values of N for each of the above populations. They are shown in Tables 5 and 6 against their corresponding normal-theory values. Differences between the mean values and the ratios of the standard deviations are also tabled. It is worth noting that with increasing sample size, the ratio of standard deviations hardly ever tends to unity. This shows that the variance of z is very sensitive to changes in the population form. In most cases of large samples the actual variance is found to be widely divergent from its normal theory value. The magnitude of the differences in mean z, however, diminishes gradually (although not very rapidly) as the sample size increases.

Table 5. *The comparative values of (mean $z-\zeta$) of the distribution of z,*

	Population	Population value of ρ	$N = 21$			$N = 51$		
			(i) Non-normal samples	(ii) Normal samples	Difference |(i) − (ii)|	(i) Non-normal samples	(ii) Normal samples	Difference |(i) − (ii)|
I	Trumps	−0·3305	−0·01089	−0·00774	0·00315	−0·00430	−0·00322	0·00108
II	Full year barometric heights	0·5807	0·02419	0·01549	0·00870	0·00979	0·00596	0·00383
III	Newborn child	0·6455	0·03029	0·01723	0·01306	0·01327	0·00663	0·00664
IV_0	Contemporaneous barometric heights*	0·7808	0·02416	0·02089	0·00327	0·00954	0·00803	0·00151
V	Beans	0·7811	0·01891	0·02090	0·00199	0·00709	0·00803	0·00094

* Note that the figures in this row are for the original population IV_0 (see p. above).

Table 6. *The comparative values of the standard deviation of the dist*

	Population	Population value of ρ	$N = 21$			$N = 51$		
			(i) Non-normal samples	(ii) Normal samples	Ratio (i)/(ii)	(i) Non-normal samples	(ii) Normal samples	Ratio (i)/(ii)
I	Trumps	−0·3305	0·23101	0·23517	0·982	0·14501	0·14424	1·005
II	Full year barometric heights	0·5807	0·25169	0·23449	1·073	0·15591	0·14408	1·082
III	Newborn child	0·6455	0·25307	0·23425	1·080	0·15744	0·14402	1·093
IV_0	Contemporaneous barometric heights*	0·7808	0·23893	0·23366	1·023	0·14698	0·14388	1·022
V	Beans	0·7811	0·28090	0·23366	1·202	0·16542	0·14388	1·150

* Note that the figures in this row are for the original population IV_0 (see p. above).

Actual frequency curves for z in samples of 11 and 21, from the population IV, as considered in § 4, have been shown in Figs. 2a, b respectively. Corresponding curves of z for the normal population have also been drawn in each case for comparison. Comparative values of the mean and of σ_z for the population IV and the normal population are given in Table 7. The values of $\gamma_2(z)$ are:

	$N = 11$	$N = 21$
Population IV	0·6475	0·2693
Normal population	0·2405	0·1101

of z, in samples from normal and non-normal populations

$N = 101$			$N = 201$			$N = 401$		
(i) Non-normal samples	(ii) Normal samples	Difference $\lvert(i)-(ii)\rvert$	(i) Non-normal samples	(ii) Normal samples	Difference $\lvert(i)-(ii)\rvert$	(i) Non-normal samples	(ii) Normal samples	Difference $\lvert(i)-(ii)\rvert$
−0·00217	−0·00163	0·00054	−0·00109	−0·00082	0·00027	−0·00055	−0·00041	0·00014
0·00491	0·00294	0·00197	0·00246	0·00146	0·00100	0·00123	0·00073	0·00050
0·00683	0·00327	0·00356	0·00346	0·00163	0·00183	0·00174	0·00081	0·00093
0·00475	0·00396	0·00079	0·00237	0·00197	0·00040	0·00118	0·00098	0·00020
0·00347	0·00396	0·00049	0·00171	0·00197	0·00026	0·00085	0·00098	0·00013

distribution of z, in samples from normal and non-normal populations

$N = 101$			$N = 201$			$N = 401$		
(i) Non-normal samples	(ii) Normal samples	Ratio (i)/(ii)	(i) Non-normal samples	(ii) Normal samples	Ratio (i)/(ii)	(i) Non-normal samples	(ii) Normal samples	Ratio (i)/(ii)
0·10236	0·10099	1·014	0·07233	0·07106	1·018	0·05113	0·05012	1·020
0·10954	0·10093	1·085	0·07722	0·07104	1·087	0·05452	0·05011	1·088
0·11083	0·10091	1·098	0·07828	0·07103	1·102	0·05525	0·05011	1·103
0·10307	0·10086	1·022	0·07260	0·07101	1·023	0·05125	0·05011	1·023
0·11404	0·10086	1·131	0·07961	0·07101	1·121	0·05592	0·05011	1·116

8. The normal approximation to the distribution of z in samples from a non-normal population

The results of the previous section show that the distribution of z, in samples from population (1), is not far from normality with mean as in (73) and variance as in (74) provided the samples are of reasonable size. Accordingly, the problem of computation of the probability integral of r, for the frequency function (32), may be put, without serious loss of accuracy, in terms of those of the normal distribution.

In the normal-theory case, the approximation

$$\text{mean}\, z = \zeta + \frac{\rho}{2(N-1)} = \bar{z}_0, \tag{81}$$

$$\text{var}\, z = \frac{1}{N-1} + \frac{4-\rho^2}{2(N-1)^2} = \Sigma_0^2(z), \tag{82}$$

considers terms in mean z, (57), up to order $(N-1)^{-1}$, and those in var z, (61), up to order $(N-1)^{-2}$. In practical application the above approximation, equations (81)–(82), is made still rougher by using

$$\text{mean}\, z = \bar{z}_0, \tag{81'}$$

$$\text{var}\, z = \frac{1}{N-3}, \tag{82'}$$

for while shortening the calculation needed, it leads in general to a sufficiently close estimate of the actual probability integral in samples of moderate size.

In the present case we may take a similar approximation, as in (81)–(82), considering terms up to order $(N-1)^{-1}$, in (73) and those up to order $(N-1)^{-2}$, in (74). Thus we have,

$$\begin{aligned}\text{mean}\, z &= \bar{z}_0 + \frac{\rho(3-\rho^2)(\lambda_{40}+\lambda_{04}) - 4(1+\rho^2)(\lambda_{31}+\lambda_{13}) + 2\rho(5+\rho^2)\lambda_{22}}{8(N-1)(1-\rho^2)^2} \\ &= \theta(z), \text{ say}\end{aligned} \tag{83}$$

$$\begin{aligned}\text{var}\, z = \Sigma_0^2(z) &+ \left\{\frac{1}{(N-1)} - \frac{2+\rho^2}{(N-1)^2}\right\}\left\{\frac{\rho^2(\lambda_{40}+\lambda_{04}) - 4\rho(\lambda_{31}+\lambda_{13}) + 2(2+\rho^2)\lambda_{22}}{4(1-\rho^2)^2}\right\} \\ &+ \frac{1}{(N-1)^2}\{\tfrac{1}{2}[-\rho^2(3-\rho^2)(\lambda_{30}^2+\lambda_{03}^2) + 4\rho(\lambda_{30}\lambda_{03}) - (4+13\rho^2+\rho^4)(\lambda_{21}^2+\lambda_{12}^2) \\ &\quad + 12\rho(2+\rho^2)(\lambda_{21}\lambda_{12}) - 4(1+2\rho^2)(\lambda_{30}\lambda_{12}+\lambda_{03}\lambda_{21}) \\ &\quad + 2\rho(5+\rho^2)(\lambda_{30}\lambda_{21}+\lambda_{03}\lambda_{12})]/(1-\rho^2)^3\} \\ &= \Sigma^2(z), \text{ say.}\end{aligned} \tag{84}$$

It may be of some interest to compare here the customary normal-theory approximations involved in the use of z with those reached by the suggested corresponding shorter expressions (83)–(84), for the mean and the variance of z, in the case of the non-normal population IV based on the frequency of contemporaneous barometric heights. Table 7 below shows these results for samples of size $N = 11$ and $N = 21$.

The approximations (83)–(84) for the non-normal case are seen to be as good as (81)–(82) in the normal case. What is most noticeable, however, is the very considerable reduction in σ_z for population IV, compared with the 'normal theory' figures.

Results reached for different approximations in the normal-theory case are given on the left-hand side of Table 8. It will be seen that in a case like $N = 11$, for the normal population ($\rho = 0{\cdot}8$) both the approximations (81)–(82) and (81')–(82') are sufficiently close to the actual

distribution of z, though the former appears to be more satisfactory than the latter. In general, the use of the actual value of ρ in the expression for the variance (i.e. the form of approximation (81)–(82)) leads, particularly in the case of small samples, to closer estimates of the probability of z.

Table 7. *Comparing approximations reached by the suggested shorter expressions (formulae (83)–(84)) for the mean and the variance of z*

Size of sample	Population IV based on the frequency of contemporaneous barometric heights				Normal population			
	Mean z from formula		σ_z from formula		Mean z from formula		σ_z from formula	
	(73)	(83)	(74)	(84)	(57)	(81)	(81)	(82)
$N = 11$	1·1398	1·1356	0·3139	0·3099	1·1443	1·1386	0·3459	0·3418
$N = 21$	1·1182	1·1171	0·2070	0·2062	1·1200	1·1186	0·2336	0·2328

Table 8. *Comparison of the probability integrals of z under various assumptions for the case* $N = 11$, $\rho = 0{\cdot}8$

r	z	Probability integral, sampling from a normal population			Probability integral, sampling from population IV	
		(i) Normal approx. $(\bar{z}_0, 1/\sqrt{(N-3)})$	(ii) Normal approx. $(\bar{z}_0, \Sigma_0(z))$	Actual, calculated from (28)	Normal approx. $(\theta(z), \Sigma(z))$	Actual, calculated from (32)
0·00	0·0000	0·00064	0·00043	0·00089	0·00013	0·00019
+0·10	0·1003	0·00166	0·00119	0·00194	0·00042	0·00056
+0·20	0·2027	0·00406	0·00309	0·00419	0·00132	0·00154
+0·30	0·3095	0·00951	0·00763	0·00900	0·00385	0·00413
+0·40	0·4236	0·02158	0·01822	0·01940	0·01080	0·01085
+0·50	0·5493	0·04778	0·04232	0·04223	0·02925	0·02830
+0·60	0·6931	0·10384	0·09621	0·09311	0·07668	0·07364
+0·650	0·7753	0·15207	0·14388	0·13872	0·12247	0·11807
+0·700	0·8673	0·22143	0·21364	0·20658	0·19328	0·18787
+0·750	0·9730	0·31970	0·31394	0·30596	0·29981	0·29451
+0·800	1·0986	0·45496	0·45341	0·44643	0·45244	0·44876
+0·825	1·1723	0·53793	0·53923	0·53348	0·54704	0·54432
+0·850	1·2562	0·63023	0·63455	0·63020	0·65129	0·64901
+0·875	1·3540	0·72883	0·73575	0·73243	0·75947	0·75666
+0·900	1·4722	0·82731	0·83550	0·83226	0·86126	0·85650
+0·925	1·6226	0·91449	0·92163	0·91764	0·94193	0·93573
+0·950	1·8318	0·97504	0·97873	0·97489	0·98765	0·98254
+0·975	2·1847	0·99846	0·99890	0·99776	0·99964	0·99876
+1·000	∞	1·00000	1·00000	1·00000	1·00000	1·00000

To examine the adequacy of the approximation (83)–(84) in small samples from moderately non-normal populations, we shall apply it in the case of the distribution of r in samples of 11 from the population IV, based on contemporaneous barometric heights, already obtained in §4.

Thus we have for $N = 11$,

$$\text{mean } z = \theta(z) = 1{\cdot}135649, \quad \sigma_z = \Sigma(z) = 0{\cdot}309923.$$

The results given on the right-hand side of Table 8 show a satisfactory agreement with the actual distribution, for the non-normal case considered. Indeed, the agreement compares well with that of the corresponding normal-theory case.

Summary

The mathematical form of the distribution of r in non-normal samples is obtained, the parent population being specified by the bivariate Edgeworth surface including terms as far as those in $\lambda_{30}^2, \lambda_{30}\lambda_{21}, \ldots, \lambda_{03}\lambda_{12}$ and λ_{03}^2. Being applicable for any ρ, the derived law holds good (i) for any size of sample if the fifth and higher order population λ's are negligible, and (ii) for any population provided the samples are fairly large; it is not unlikely that the formula has quite an extended range of applicability for moderate size of samples.

The tables of the ordinates of the corrective functions due to fourth-order and the square and product of third-order semi-invariants (measuring respectively the excess and skewness of the population) are obtained for $N = 11, \rho = 0{\cdot}0$ and $0{\cdot}8$ and $N = 21, \rho = 0{\cdot}8$. Frequency curves for r and z (Fisher's logarithmic transformation of the coefficient of correlation) in samples of 11 and 21, from a moderately non-normal population ($\rho = 0{\cdot}8$ approximately) are drawn for comparison with the corresponding curves of the normal-theory case.

In cases where ρ is not zero, the normal-theory law of r is found to be affected by changes in the population form however large the sample size is. But when ρ is zero the effect of non-normality on the distribution of r is not very serious and is smaller still if the variables are completely independent. The true probabilities of r may be examined from the derived distribution for *a priori* values of ρ and the λ's of the population.

The transformed variate z, also, is considerably influenced by the non-normality of the population if ρ is not zero, but the disturbance is largely in the mean and variance, the distribution being still nearly normal (although the approach towards normality is not so rapid as in the normal-theory case). Practical examples taken indicate that for samples of reasonable size the assumption of the normality of z is a remarkably good approximation. Accordingly, the problem of computation of the probability integral of r for the derived law may be put in terms of those of the normal distribution without serious loss of accuracy. Illustration shows that for samples as small as 11 from a moderately non-normal population, it is sufficient to include in mean z, (73), terms up to $O[(N-1)^{-1}]$ and in var z, (74), those up to $O[(N-1)^{-2}]$ only, thereby shortening the calculation of the probability integral of r considerably.

In conclusion, I should like to acknowledge my indebtedness to Dr H. E. Daniels for his kind advice and criticism in the course of my investigations; also to Prof. E. S. Pearson for suggesting a number of improvements to the paper.

REFERENCES*

DAVID, F. N. (1938). *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples.* Cambridge University Press.

FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**, 507.

FISHER, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, no. 4, 1.

KENDALL, M. G. (1949). Rank and Product-moment correlation. *Biometrika*, **36**, 177.

PEARSON, E. S. (1929). Some notes on sampling tests with two variables. *Biometrika*, **21**, 337.

PEARSON, K. (1925). The fifteen constant bivariate frequency surface. *Biometrika*, **17**, 268.

PEPPER, J. (1929). Studies in the theory of sampling. *Biometrika*, **21**, 231.

PRETORIUS, S. J. (1930). Skew bivariate frequency surfaces examined in the light of numerical illustrations. *Biometrika*, **22**, 109.

QUENSEL, C. E. (1938). The distributions of the second moment and of the correlation coefficient in samples from populations of Type A. *Lunds. Univ. Arss.* N.F. Adv. 2, Bd. 34, Nr. 4; *Kungl. Fys. Sal. Handl.* N.F. Bd. 49, Nr. 4.

RIDER, P. R. (1932). The distribution of the correlation coefficient in small samples. *Biometrika*, **24**, 382.

ROMANOVSKY, V. (1925). On the moments of standard deviations and of correlation coefficient in samples from normal population. *Metron*, **5**, no. 4, 3.

SOPER, H. E., YOUNG, A. W., CAVE, B. M., LEE, A. & PEARSON, K. (1917). A Co-operative Study—On the distribution of the correlation coefficient in small samples. *Biometrika*, **11**, 328.

TSCHUPROW, A. A. (1925). *Grundbegriffe und Grundprobleme der Korrelationstheorie.* Leipzig: Teubner (English translation as the *Mathematical Theory of Correlation*, William Hodge and Co. Ltd., 1939).

VAN UVEN, M. J. (1925–9). On treating skew correlation. *Proc. K. Akad. Wet. Amst.* **28**, nos. 8–9, 797; **28**, no. 10, 919; **29**, no. 4, 580; **32**, no. 4, 408.

WICKSELL, S. D. (1917). The correlation function of Type A and the regression of its characteristics. *K. Svenska VetenskAkad. Handl.* Bd. **58**, no. 3, 1–48.

* A bibliography of the chief papers dealing with artificial sampling experiments from non-normal populations is given by G. B. Hey, *Biometrika*, **30**, 79 (1938).

MISCELLANEA

Some observations on the practical aspects of weighing designs

By K. S. BANERJEE, *Pusa, Bihar, India*

1. Introduction

Certain numerical illustrations of the author's methods were given in a previous paper (Banerjee, 1950), and the following notes are intended to assist in understanding the working, and also to discuss some of the practical aspects of weighing designs.

2. Spring balance designs in balanced incomplete blocks

Let v objects be weighed in combinations of k at a time ($k<v$), such that each pair of objects occur together in the same number (λ) of weighings. Let r be the number of times each object is included in a weighing (number of replicates) and b be the number of weighings (blocks). Then, as is well known from the theory of balanced incomplete blocks, $rv = bk$ and $(v-1)\lambda = r(k-1)$. Suppose that the records of the successive weighings are $y_1, y_2, \ldots, y_b$. Let $w_1, w_2, \ldots, w_v$ be the estimates (by least squares) of the weights $\omega_1, \omega_2, \ldots, \omega_v$ of the v objects. Then the normal equations determining the w_i are in general (1950, pp. 50–1)

$$\left.\begin{aligned} rw_1+\lambda w_2+\lambda w_3+\ldots+\lambda w_v &= z_1\\ \lambda w_1+rw_2+\lambda w_3+\ldots+\lambda w_v &= z_2\\ \ldots\ldots\ldots\

Also
$$\sum_1^b (y_j - Y_j)^2 = \sum_1^b y_j(y_j - Y_j)$$
$$= \sum_1^b y_j^2 - \sum_1^b y_j Y_j$$
$$= \sum_1^b y_j^2 - \sum_1^v w_i z_i$$

from standard least square theory. This is a convenient way of calculating the sum of the squared residuals. Another is to calculate
$$\sum_1^b (y_j - \bar{y})^2 - \frac{1}{r-\lambda} \sum_1^v (z_i - \bar{z})^2,^*$$

where $\bar{y}$ and $\bar{z}$ denote the means of y_j and z_i respectively.

Then the estimated variance of y will be
$$s_y^2 = \frac{\sum_1^b (y_j - Y_j)^2}{b-v}$$

from which the estimated variance of w_i can be obtained. In the above illustration, for example,
$$s_{w_i}^2 = \tfrac{7}{16} s_y^2,$$

being the same for all w_i. It should be noted, however, that no estimate is possible with this illustration, nor, indeed, with any other b.i.b. design in which $b = v$. For in such cases we have $Y_j = y_j$ for all j and thus
$$\sum_1^b (y_j - Y_j)^2 = 0.$$

In the general case $s_{w_i}^2$ will be s_y^2 multiplied by
$$\frac{1}{r} + \frac{\lambda^2(b-r)}{r^2(r-\lambda)^2}$$

for b.i.b. designs.

3. The elimination of a bias suspected after the weighing operations have been completed in a spring balance

It has been shown (Hotelling, 1944; Mood, 1946) how it is possible to adjust the design matrix to suit the estimation of the weights of the objects in a biased chemical balance. The device is to reduce the elements of the first column of the design matrix to $+1$ (1944, 1946) without affecting the nature and efficiency of the design. But in the case of a spring balance design, where the elements of the matrix are restricted to be $+1$ or 0, it is not possible to take account of the bias by reducing the elements of the first column of the design matrix to $+1$ without affecting the design. If, instead, an additional column of $+1$'s is added to the design, corresponding to a bias ω_0, which can be looked on as the weight of an additional object, and the matrix is used in the weighing operations with the biased balance; then the normal equations will be $(v+1)$ in number, as follows:

$$\left.\begin{array}{lll} bw_0 + rw_1 + rw_2 & & = t' \\ rw_0 + & & = z_1' \\ rw_0 + & \text{(The left-hand side} & = z_2' \\ \vdots & \text{of equations (1))} & \vdots \\ rw_0 + & & = z_v' \end{array}\right\}, \qquad (2)$$

where $t', z_1', z_2', \ldots, z_v'$ have the same significance as before in relation to the new records of weighings with the biased balance. Adding up both sides of (2), we have
$$(b+rv)w_0 + \{2r + \lambda(v-1)\}w_1 + \{2r + \lambda(v-1)\}w_2 + \ldots + \{2r + \lambda(v-1)\}w_v = t' + kt',$$

i.e.
$$(k+1)(bw_0 + rw_1 + rw_2 + \ldots + rw_v) = (k+1)t'.$$

Cancelling $(k+1)$ from both sides, we get the first equation of (2). The equations (2) are not therefore independent, and as such, the estimate of ω_0 cannot be obtained. If, on the other hand, a column of ones and a row of zeros in that order are added to the design matrix (1950, p. 52), it will be possible to estimate the bias. The normal equations in such a situation will take the form of (2) except for the first equation, which will be
$$(b+1)w_0 + rw_1 + rw_2 + \ldots + rw_v = y_0 + t',$$

* This result was communicated to me by Dr John Wishart.

where y_0 is the record of weighing with empty pans, i.e. the record of weighing corresponding to the bias. Adding up the equations and subtracting the total from the first, we get $w_0 = y_0$. Subtracting ry_0 from all the equations beginning from the second, we get a new set of equations similar in form to (1), which can be easily solved as before.

It now remains to see if it is possible to eliminate a suspected bias from the weighing operations, the bias being suspected after the weighing operations have already been completed. In such a situation, the value of y_0 will be unknown.

Let us refer to the design matrix, where an additional column of ones and a row of zeros are added to a b.i.b. design to suit a biased balance, and let us suppose that the weighing operations have been made under the assumption that the balance was free of bias. Denoting by w_i the estimate of the ith object when no bias is suspected and by w_i^* the estimate when a bias is suspected, we get, by solving the normal equations,

$$\begin{aligned} w_i^* &= \frac{1}{(r-\lambda)}\left\{(z_i - ry_0) - \frac{\lambda}{r+\lambda(v-1)}(T - rvy_0)\right\} \\ &= \frac{1}{(r-\lambda)}\left(z_i - \frac{\lambda T}{rk}\right) - \frac{(kr-\lambda v)}{(r-\lambda)k}y_0 \\ &= w_i - \frac{1}{k}y_0. \end{aligned}$$

If now an attempt is made to estimate y_0 by minimizing the error variance, as is commonly done under the missing plot technique in agrobiological experiments (Yates, 1933), it is noticed that y_0 does not occur in the expression for the error variance on account of the fact that in each weighing operation, the objects are weighed k at a time. While stating that the estimates cannot be determined free of the bias, it is noticed that the sum of squares of the residuals does not undergo any change.

The elimination of the bias may however be possible, if the design matrix is of the sub-matrix type or design incomplete in number of columns. An illustration (Mood, 1946) of seven columns of $L_{11}(r = 6, \lambda = 3)$, with the records of weighings, is furnished below:

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7		g.
1	1	0	1	0	0	0	1	=	9·11926
2	1	1	0	1	0	0	0	=	11·33376
3	0	1	1	0	1	0	0	=	6·95805
4	1	0	1	1	0	1	0	=	10·63964
5	1	1	0	1	1	0	1	=	14·45202
6	1	1	1	0	1	1	0	=	13·86458
7	0	1	1	1	0	1	1	=	9·78492
8	0	0	1	1	1	0	1	=	6·85182
9	0	0	0	1	1	1	0	=	4·63865
10	1	0	0	0	1	1	1	=	10·02561
11	0	1	0	0	0	1	1	=	6·05214

Each object is repeated an equal number of times (r), as also each pair of objects (λ), but the number of objects in the different weighing operations is not constant. Solving the normal equations, which are of the same nature as in (1), the estimates are obtained as 5·85823, 3·52843, 1·78603, 1·94687, 1·64352, 1·04845 and 1·47520. It may be noted in particular that in such designs

$$\sum_1^b (y_j - Y_j) \neq 0.$$

Let us suppose that the above design has been used in the weighing operations with a biased balance and that the value y_0 is not known. Solving the normal equations, the estimate of the bias is obtained as

$$w_0 = \tfrac{2}{3}y_0 + \tfrac{1}{3}(0{\cdot}00017). \quad \text{Here } w_i^* = w_i - \tfrac{1}{7}w_0.$$

Solving the equation obtained by minimizing the error variance with respect to y_0 (1933), the value of y_0, as also that of w_0, is obtained as 0·00017. The amount of the necessary adjustment is therefore given by $\frac{1}{7}w_0 = 0{\cdot}00004$.

4. Standardization of weight-units

As standardization also means weighing operations, repeated weighings are needed to get steady estimates. An illustration is furnished here to show that the standardized weight-units may perhaps be obtained without making any separate weighings. The procedure may be adopted in situations where

the weight-units to be standardized are not too numerous to render the solution of the normal equations lengthy.

Let us refer to the b.i.b. design given by $r = 5$, $b = 10$, $v = 6$, $k = 3$, $\lambda = 2$, and let us suppose that it was necessary to standardize the largest weight-units, namely the weight-units of 5 and 10 g. only. Two columns may be added to the design matrix to represent these two adjustments supposed as two additional objects (ω_{00} and ω_{01}) to be weighed, the two columns being filled in with 1 or 0 as the case may be. For example, if, of the two weight-units which require standardization, the weight-unit of 5 g. only is needed, among other smaller weight-units, in a weighing operation, the weighing should be completed as usual, inserting a plus one in the column for ω_{00} and a zero in the column for ω_{01} in the relevant row. If neither of the two weight-units is used in any weighing operation, there will be zeros under both the columns in that particular row. The record of weighings with the modified design matrix is shown below:

ω_{00}	ω_{01}	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6		g.
0	1	1	1	1	0	0	0	=	11·17575
1	0	0	1	1	0	0	1	=	6·36704
0	1	1	1	0	1	0	0	=	11·33746
1	0	0	1	0	1	1	0	=	7·12503
1	0	1	0	1	0	1	0	=	9·29941
1	0	0	1	0	0	1	1	=	6·22389
1	0	1	0	0	1	0	1	=	8·86387
1	0	0	0	1	1	1	0	=	5·37762
1	0	1	0	0	0	1	1	=	8·56015
0	0	0	0	1	1	0	1	=	4·79442

The normal equations are given by:

w_{00}	w_{01}	w_1	w_2	w_3	w_4	w_5	w_6		g.	
7	0	3	3	3	3	5	4	=	51·81701	(1)
0	2	2	2	1	1	0	0	=	22·51321	(2)
3	2	5	2	2	2	2	2	=	47·23664	(3)
3	2	2	5	2	2	2	2	=	42·22917	(4)
3	1	2	2	5	2	2	2	=	37·01424	(5)
3	1	2	2	2	5	2	2	=	37·49840	(6)
5	0	2	2	2	2	5	2	=	36·58610	(7)
4	0	2	2	2	2	2	5	=	34·80937	(8)

The solution of the normal equations is not difficult. If $\frac{2}{15}$ of the sum of equations (3) to (8) (i.e. the equations corresponding to $w_1, w_2, \ldots, w_6$) be subtracted from each, the left-hand sides will contain only the variables w_{00} and w_{01} and an additional variable w_i. These equations may be utilized to eliminate the w_i's from the first two equations. In the present instance, the equations in w_{00} and w_{01} are obtained as

$$14w_{00} - 6w_{01} = -0{\cdot}026610,$$

$$-6w_{00} + 4w_{01} = -0{\cdot}027480.$$

When the values of w_{00} and w_{01} are known, appropriate multiples of these values may be subtracted from each of the equations (3) to (8) to get a set of equations which admit of easy solution. The estimates of the adjustments, w_{00} and w_{01}, and of the objects, are obtained as $-0{\cdot}01357$, $-0{\cdot}02722$; and 5·87405, 3·53823, 1·79085, 1·95223, 1·64811 and 1·05134 respectively.

My grateful thanks are due to Prof. R. C. Bose and to Dr John Wishart for helpful suggestions in connexion with this work.

REFERENCES

Banerjee, K. S. (1950). How balanced incomplete block designs may be made to furnish orthogonal estimates in weighing designs. *Biometrika*, **37**, 50.

Hotelling, H. (1944). Some improvements in weighing and other experimental techniques. *Ann. Math. Statist.* **15**, 297.

Mood, A. M. (1946). On Hotelling's weighing problem. *Ann. Math. Statist.* **17**, 432.

Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Emp. J. Exp. Agric.* **1**, 129.

Test for the significance of the difference between means in two normal populations having unequal variances

By D. G. C. GRONOW, *University College, London*

It is assumed that samples of size n_1 and n_2 are drawn from normal populations Π_1 and Π_2 respectively. The sample units are x_{ti} ($t = 1, 2$; $i = 1, 2, \ldots, n_t$) and the question is posed: 'Are the data consistent with the hypothesis that the two population means ξ_1 and ξ_2 are equal?' In the case where the population variances cannot be assumed equal, several criteria have been put forward to test the hypothesis $\xi_1 = \xi_2$, and the problem has been discussed by numerous writers. In the present paper we shall consider only two test statistics which Welch (1937) denoted by u and v:

$$u = \frac{\bar{x}_1 - \bar{x}_2}{\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left(\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\right)\right]^{\frac{1}{2}}}, \quad v = \frac{\bar{x}_1 - \bar{x}_2}{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^{\frac{1}{2}}},$$

where

$$s_i^2 = \sum_{i=1}^{n_i} (x_{ti}-\bar{x}_i)^2/(n_i-1).$$

It seemed of interest to investigate the power of u and v to establish significance when there is a real difference in the population means and the variances may be unequal. This problem has already been solved in principle by Hsu (1938), but his solution is not in a form which readily allows of numerical calculation.

The method of procedure followed has been to obtain approximations to the moments of the distributions of u and v by the use of Fisher's k-statistics. Thus in the general approach to the problem, of which the work here is a simplified case, nothing is assumed about the two populations except that the cumulants of each exist, but are different one from the other. In the simplified case considered in this note, we suppose that the two populations are normal but may have different means and variances. Giving dashes to the k-statistics of the second population sampled, it is clear that both u and v may be written in the form

$$b(k_1 - k_1')(k_2 + ak_2')^{-\frac{1}{2}},$$

where the constants a and b are different for u and for v, except where $n_1 = n_2$, when $u \equiv v$. A single expansion will thus serve for both cases.

Let

$$z = (k_1 - k_1')(k_2 + ak_2')^{-\frac{1}{2}}.$$

Expanding by Taylor's theorem about the point $(\kappa_1, \kappa_1', \kappa_2, \kappa_2')$, taking expectations and collecting terms, we have, to order $(n-1)^{-4}$,

$$\begin{aligned}
\mathscr{E}(z) = \frac{\delta}{\tau}\Bigg[& 1 + \frac{1}{n_1-1}\frac{3\kappa_2^2}{4\tau^4} + \frac{1}{n_2-1}\frac{3a^2\kappa_2'^2}{4\tau^4} \\
& + \frac{1}{(n_1-1)^2}\frac{5\kappa_2^3}{2\tau^6}\left(-1+\frac{21\kappa_2}{16\tau^2}\right) + \frac{1}{(n_1-1)(n_2-1)}\frac{105a^2\kappa_2^2\kappa_2'^2}{16\tau^8} + \frac{1}{(n_2-1)^2}\frac{5a^3\kappa_2'^3}{2\tau^6}\left(-1+\frac{21a\kappa_2'}{16\tau^2}\right) \\
& + \frac{1}{(n_1-1)^3}\frac{105\kappa_2^4}{8\tau^8}\left(1-\frac{3\kappa_2}{\tau^2}+\frac{33\kappa_2^2}{16\tau^4}\right) + \frac{1}{(n_1-1)^2(n_2-1)}\frac{315a^2\kappa_2^3\kappa_2'^2}{8\tau^{10}}\left(-1+\frac{33\kappa_2}{16\tau^2}\right) \\
& + \frac{1}{(n_1-1)(n_2-1)^2}\frac{315a^2\kappa_2^2\kappa_2'^3}{8\tau^{10}}\left(-1+\frac{33a\kappa_2'}{16\tau^2}\right) + \frac{1}{(n_2-1)^3}\frac{105a^4\kappa_2'^4}{8\tau^8}\left(1-\frac{3a\kappa_2'}{\tau^2}+\frac{33a^2\kappa_2'^2}{16\tau^4}\right) \\
& + \frac{1}{(n_1-1)^4}\frac{21\kappa_2^5}{2\tau^{10}}\left(-9+\frac{715\kappa_2}{16\tau^2}-\frac{2145\kappa_2^2}{32\tau^4}+\frac{32175\kappa_2^3}{1024\tau^6}\right) \\
& + \frac{1}{(n_1-1)^3(n_2-1)}\frac{3465a^2\kappa_2^4\kappa_2'^2}{32\tau^{12}}\left(3-\frac{13\kappa_2}{\tau^2}+\frac{195\kappa_2^2}{16\tau^4}\right) \\
& + \frac{1}{(n_1-1)^2(n_2-1)^2}\frac{1155a^3\kappa_2^3\kappa_2'^3}{4\tau^{12}}\left(1-\frac{39(\kappa_2+a\kappa_2')}{16\tau^2}+\frac{1755a\kappa_2\kappa_2'}{256\tau^4}\right) \\
& + \frac{1}{(n_1-1)(n_2-1)^3}\frac{3465a^4\kappa_2^2\kappa_2'^4}{32\tau^{12}}\left(3-\frac{13a\kappa_2'}{\tau^2}+\frac{195a^2\kappa_2'^2}{16\tau^4}\right)
\end{aligned}$$

$$+\frac{1}{(n_2-1)^4}\frac{21a^5\kappa_2'^5}{2\tau^{10}}\left(-9+\frac{715a\kappa_2'}{16\tau^2}-\frac{2145a^2\kappa_2'^2}{32\tau^4}+\frac{32175a^3\kappa_2'^3}{1024\tau^6}\right)\Bigg],$$

where $\delta = \kappa_1 - \kappa_1' = \xi_1 - \xi_2$ and $\tau = (\kappa_2 + a\kappa_2')^{\frac{1}{2}}$.

By taking second, third and fourth powers of z, expanding in a way similar to the above and taking expectations, the second, third and fourth moments of z were obtained. These moments are, however, of great length, and there appears to be little point in reproducing them here. Numerical values were substituted in these expressions, and the appropriate Pearson or Gram-Charlier Type A curve (with the same mean, standard deviation, β_1 and β_2) was assumed to be an approximation to the unknown true distribution.

Following the procedure briefly outlined, the distributions of u and v under hypothesis H_0 (i.e. $\delta = 0$) were first investigated. As Welch (1937) has already pointed out, if the variances are not the same in the two populations sampled, the assumption that u is distributed as 'Student's' t with $n_1 + n_2 - 2$ degrees of freedom will lead to bias in tests of significance. His results are confirmed by the present investigation. Taking the case of $n_1 = n_2 = 10$ ($u = v$), a Pearson Type VII with mean at zero and second and fourth moments obtained by the k-statistic method was used as an approximation to the true distribution. By this means, the probability that $u\ (= v)$ falls beyond the 5 and 1 % significance levels of the t-distribution with 18 degrees of freedom was calculated. These approximate probabilities are shown in Table 1*a*, the double-tailed test having been considered.

Table 1*a*. *Probability, as a percentage, that* $u\ (= v)$ *falls beyond the* 5 *and* 1 % *points of the t-distribution with* 18 *degrees of freedom, calculated by the approximate method*

Case: $n_1 = n_2 = 10$, $\delta = 0$

Ratio of → κ_2'/κ_2	1	1·5	2	2·5	3
5 %	5·00	5·05	5·13	5·22	5·30
1 %	1·00	1·01	1·04	1·09	1·13

The figures in the 5 % row correspond to those which can be read from the curve (*a*) given in Fig. 1 of Welch's (1937) paper, where he used a different method of approximation. The bias in this case is nowhere very serious.

Similar calculations for the case $n_1 = 15$, $n_2 = 5$, $\delta = 0$ were carried out. These are given in Table 1*b*, the results in this case being different for u and for v.

Table 1*b*. *Probability, as a percentage, that* u *and* v *fall beyond the* 5 *and* 1 % *points of the t-distribution with* 18 *degrees of freedom*

Case: $n_1 = 15$, $n_2 = 5$, $\delta = 0$

Criterion	Ratio of → κ_2'/κ_2	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3
u	5 %	1·40	2·25	5·00	9·77	13·20
	1 %	0·15	0·31	1·00	2·81	4·55
v	5 %	5·47	5·85	6·83	7·99	8·67
	1 %	1·15	1·33	1·81	2·54	2·94

The values for u correspond approximately to those which can be read from Welch's curve (b) and those for v from his curve (c). The bias, as might be expected for the case of unequal sample sizes, is much worse that in the case previously considered. As Welch pointed out, the bias is rather less using v than using u.

The approximations to the true moments of u and v can also be used to answer the following questions. How will inequality in the population variances affect the chance of establishing the significance of a difference, δ, in means if (i) u and (ii) v are referred to the t-distribution with n_1+n_2-2 degrees of freedom? If inequality in variances is suspected, which test would it be preferable to use? As the figures in Tables 1*a* and 1*b* show, as soon as the variances become unequal, the percentage levels for t are no longer correct for u, and even when $\kappa_2 = \kappa_2'$, they are not correct for v unless $n_1 = n_2$. However, without the knowledge of κ_2'/κ_2, it is not possible to correct the levels. Let us consider, therefore, what happens if we do use the t levels, lacking any means of correction. The results of following this procedure are shown in Tables 2 and 3.

Table 2. *Power of the u and v tests to detect a difference δ between population means (referred to 5 %, t_{18} significance point)*

(*a*) Case: $n_1 = n_2 = 10$, $\alpha = 0{\cdot}05$ (error of first kind, i.e. when $\delta = 0$, $\kappa_2 = \kappa_2'$). Power of u ($= v$, in this case)

κ_2'/κ_2 \ δ/κ_2	0	0·5	1	1·5	2	2·5	3
1	0·050	0·182	0·562	0·892	0·987	0·998	1·000
1·5	0·050	0·156	0·471	0·816	0·966	0·994	0·999
2	0·051	0·139	0·406	0·743	0·934	0·989	0·998
2·5	0·052	0·127	0·357	0·675	0·896	0·977	0·996
3	0·053	0·120	0·320	0·616	0·854	0·961	0·992

(*b*) Case: $n_1 = 15$, $n_2 = 5$, $\alpha = 0{\cdot}05$. Power of u test

κ_2'/κ_2 \ δ/κ_2	0	0·5	1	1·5	2	2·5	3
$\frac{1}{3}$	0·014	0·111	0·518	0·912	0·988	0·998	1·000
$\frac{1}{2}$	0·023	0·124	0·492	0·876	0·970	0·996	0·999
1	0·050	0·148	0·446	0·792	0·958	0·994	0·998
2	0·098	0·179	0·396	0·676	0·878	0·968	0·994
3	0·132	0·194	0·367	0·603	0·805	0·926	0·977

(*c*) Case: $n_1 = 15$, $n_2 = 5$, $\alpha = 0{\cdot}05$. Power of v test

κ_2'/κ_2 \ δ/κ_2	0	0·5	1	1·5	2	2·5	3
$\frac{1}{3}$	0·055	0·259	0·744	0·968	0·994	0·999	1·000
$\frac{1}{2}$	0·059	0·220	0·653	0·935	0·987	0·997	1·000
1	0·068	0·171	0·471	0·797	0·947	0·991	0·998
2	0·080	0·132	0·299	0·594	0·829	0·940	0·978
3	0·087	0·118	0·217	0·438	0·679	0·843	0·930

Table 3. *Power of the u and v tests to detect a difference δ between population means (referred to 1 %, t_{18} significance point)*

(*a*) Case: $n_1 = n_2 = 10$, $\alpha = 0{\cdot}01$ (error of first kind, i.e. when $\delta = 0$, $\kappa_2 = \kappa_2'$). Power of u (= v, in this case)

κ_2'/κ_2 \ δ/κ_2	0	0·5	1	1·5	2	2·5	3
1	0·010	0·061	0·276	0·687	0·933	0·992	0·999
1·5	0·010	0·051	0·219	0·561	0·860	0·973	0·996
2	0·010	0·044	0·177	0·464	0·778	0·942	0·989
2·5	0·011	0·040	0·150	0·391	0·697	0·898	0·975
3	0·011	0·038	0·131	0·335	0·623	0·849	0·955

(*b*) Case: $n_1 = 15$, $n_2 = 5$, $\alpha = 0{\cdot}01$. Power of u test

κ_2'/κ_2 \ δ/κ_2	0	0·5	1	1·5	2	2·5	3
⅓	0·001	0·028	0·192	0·657	0·941	0·995	1·000
½	0·003	0·033	0·200	0·609	0·912	0·988	0·998
1	0·010	0·047	0·201	0·525	0·836	0·965	0·994
2	0·028	0·068	0·19[illegible]	0·433	0·703	0·888	0·969
3	0·045	0·082	0·191	0·382	0·613	0·808	0·923

(*c*) Case: $n_1 = 15$, $n_2 = 5$, $\alpha = 0{\cdot}01$. Power of v test

κ_2'/κ_2 \ δ/κ_2	0	0·5	1	1·5	2	2·5	3
⅓	0·011	0·101	0·473	0·878	0·983	0·994	0·999
½	0·013	0·083	0·372	0·787	0·961	0·989	0·995
1	0·018	0·071	0·232	0·557	0·827	0·947	0·988
2	0·025	0·057	0·127	0·313	0·597	0·813	0·922
3	0·029	0·044	0·098	0·195	0·402	0·627	0·789

When $n_1 = n_2 = 10$, the chance of rejecting the null hypothesis when it is true, i.e. when $\delta = 0$, never differs much from the nominal 0·05 and 0·01. Inequality in variances does, however, somewhat reduce the power of the test. For example, the chance of establishing significance at the 5 % level when $\delta = 1{\cdot}5\kappa_2$ is 0·89 if $\kappa_2' = \kappa_2$, but has fallen to 0·62 if $\kappa_2' = 3\kappa_2$.

When the sample sizes are unequal, the position is rather different. We have to take note of which sample comes from the population with the larger variance; the tables, therefore, show results for κ_2'/κ_2 greater and less than unity. Comparing Tables 2*b* and *c*, 3*b* and *c*, the following points may be noted:

(i) The significance levels of 'Student's' t may be very inaccurate; if the variances are nearly equal, the bias is less for u, but when they differ widely v is affected less than u.

(ii) Whether u or v is employed, for a given difference in means of detectable size, the chance of establishing significance is considerably larger when $\kappa_2'/\kappa_2 < 1$ than when it is > 1. This means, as one would expect, that there is a considerable gain if the larger sample is taken from the population with the larger variance.

(iii) Owing to the varying bias in the significance levels, it is not possible to make precise comparisons, but the following figures taken from Table 2 are illustrative of what occurs. While figures of this kind indicate the performance characteristic of the two procedures, it is difficult to suggest any simple basis on which to decide which it is preferable to adopt.

Table 4. *Examples of the consequences of rejecting the null hypothesis whenever* u (*or* v) *exceeds the* 5 % *level of* t *with* 18 *degrees of freedom*

κ_2'/κ_2	Test statistic	Sample sizes →	$n_1 = 15$ $n_2 = 5$	$n_1 = 10$ $n_2 = 10$	$n_1 = 5$ $n_2 = 15$
1	u	Chance of: Rejecting H_0 when $\delta = 0$ Establishing significance if $\delta = 1{\cdot}5\kappa_2$	0·050 0·792	0·050 0·892	0·050 0·792
	v	Chance of: Rejecting H_0 when $\delta = 0$ Establishing significance if $\delta = 1{\cdot}5\kappa_2$	0·068 0·797	0·050 0·892	0·068 0·797
$\frac{1}{2}$	u	Chance of: Rejecting H_0 when $\delta = 0$ Establishing significance if $\delta = 1{\cdot}5\kappa_2$	0·023 0·876	0·051 0·743	0·098 0·676
	v	Chance of: Rejecting H_0 when $\delta = 0$ Establishing significance if $\delta = 1{\cdot}5\kappa_2$	0·059 0·935	0·051 0·743	0·080 0·594
$\frac{1}{3}$	u	Chance of: Rejecting H_0 when $\delta = 0$ Establishing significance if $\delta = 1{\cdot}5\kappa_2$	0·014 0·912	0·053 0·616	0·132 0·603
	v	Chance of: Rejecting H_0 when $\delta = 0$ Establishing significance if $\delta = 1{\cdot}5\kappa_2$	0·055 0·968	0·053 0·616	0·087 0·438

Finally, it should be noted that it is to overcome these difficulties that quite different alternative test procedures have been suggested, leading to Sukhatme's (1938) tables based on the Fisher-Behrens test and Aspin's (1949) tables based on Welch's (1947) test.

REFERENCES

Aspin, Alice A. (1949). *Biometrika*, **36**, 290.
Hsu, P. L. (1938). *Statist. Res. Mem.* **2**, 1.
Sukhatme, P. V. (1938). *Sankhyā*, **4**, 39.
Welch, B. L. (1937). *Biometrika*, **29**, 350.
Welch, B. L. (1947). *Biometrika*, **34**, 28.

REVIEWS

Contributions to Mathematical Statistics. By R. A. FISHER. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1950. Price £3.

This latest of Wiley's series of publications in statistics contains forty-two of R. A. Fisher's statistical papers covering the years 1920–43, reproduced to scale from the originals by photo-offset process. Although even with this method of reproduction, the price of the volume is heavy, and may be beyond the reach of the average student, it will now be possible for all these papers to be available in Libraries, a welcome result. The book has a preface by W. A. Shewhart, the general editor of the series, and this is followed by P. C. Mahalanobis's biographical note written on the occasion of Fisher's visit to India in 1938. Each individual paper is preceded by an expository note written by the author.

In turning over these pages, statisticians of the reviewer's generation must be conscious of a wealth of crowding impressions. The history of the modern theory and practice of mathematical statistics may be said to have opened sixty years ago. The place that Karl Pearson's papers played during the period 1890–1915 was played by R. A. Fisher's papers appearing in this volume during the years 1920–40. Both series have had a profound influence in shaping the outlook of their statistical contemporaries, through what is written for all to read and through those suggestive flashes of illumination which lead the recipient forward either in agreement or disagreement.

Fisher's work started from the basis of statistical theory and practice as he found it, the theory of errors, Pearson's frequency curves, the methods of correlation and regression measurement, the application of χ^2 in many directions, the attack on the intriguing problems of heredity and evolution. The advances which he made have been very great, but the continuity in development is clear and many of the earlier papers in these *Contributions to Mathematical Statistics* are further steps along paths mapped out by his predecessor in the Royal Society Series, *Mathematical Contributions to the Theory of Evolution.*

Several of Fisher's early papers are not included in this volume; thus two papers on correlation (1921, 1924) and a third on 'Student's' t and allied distributions (1925) from *Metron* are missing, as well as the fundamental paper from *Biometrika* (1915) on the distribution of r in samples from a normal population.* The decision to omit the 1915 paper will be generally regretted, for we are deprived of the first introduction to the geometrical representation of $\bar{x}$, s, t and r, that 'exceedingly beautiful interpretation in generalised space', to use its author's own words, that must have fascinated so many students of statistics through thirty-five years.

The leading paper (**2**, 1920) in the series is one rather little known by statisticians because of its publication in the *Monthly Notices of the Royal Astronomical Society*, but it opened the way into a field crowded with new ideas which at the time no one but the author himself was ready to follow up. In this paper Fisher takes up the old problem of the astronomers, which Gauss had considered and, unknown to statisticians, Helmert had solved up to a point, that of the choice between the root-mean-square estimate (called σ_2) and the mean-deviation estimate (σ_1) of a population standard deviation (σ). Using the geometrical methods which he had put forward in 1915, Fisher derives and compares the standard errors of σ_1 and σ_2 as Helmert had done in 1875. But he goes beyond this by considering the joint distribution of σ_1, σ_2 and thence shows (*a*) that for given σ_2, the distribution of σ_1 must be independent of σ, (*b*) that this result will hold when any other estimator σ_3 is compared with σ_2. Further, (*c*) he suggests that while this property is possessed by σ_2 in sampling from the normal population, for a leptokurtic distribution such as the double exponential the roles may be reversed, so that the mean deviation, σ_1, will have this optimum property. Thus the fundamental conception of a 'sufficient statistic' begins to take shape.

Three of the early papers (**5**, 1922; **7**, 1923; and **8**, 1924) deal with the distribution of χ^2 in tests involving the comparison of discrete frequencies with their expectations. Fisher was here fighting a battle to show that the introduction of the concept of degrees of freedom would straighten out inconsistencies which it was beginning to be seen must follow from Pearson's formulation of the χ^2 test. He won his point, but was handicapped, as all who try to tackle these problems must be, by the impossibility of reaching any final results in neat mathematical terms because of the discontinuous character of the multinomial distribution. His style of presentation, too, did not perhaps make it easy for his elders to grasp what he was after. The reviewer was particularly interested to see paper (**7**), which he does not think he had ever seen before,

* [In view of the author's comment on the reason for the absence of the *Metron* papers, it should be made clear that permission to reproduce this paper from *Biometrika* would of course have been given had it been asked for. ED.]

probably because of its publication in *Economica*. It contains perhaps the clearest exposition of the three, and would certainly have helped him in puzzling out the problem at the time, had he read it. One is glad to see paper (**4**, 1922) made available from the *Annals of Applied Biology*, giving a demonstration, though perhaps not what would be regarded as a mathematical proof, that the index of dispersion for samples from a Poisson distribution, $\Sigma(x_i - \bar{x})^2/\bar{x}$, is distributed as χ^2.

Papers (**10**, 1922) and (**11**, 1925) contain Fisher's fundamental work on his theory of estimation; to him they formed the basis for the later development of his theory of inductive inference, continued in nos. (**24**, 1934) and (**26**, 1935) and through the series of papers on fiducial probability (**22**, 1930), (**25**, 1935), (**27**, 1936) and (**35**, 1939). We are carried into a field where the construct of a mathematical theory must be linked with the processes of rational thought. Fisher's work in this category has been fruitful, in part directly and in part by the challenge thrown out, which has forced statisticians to think and where they disagree to clarify and formulate their own ideas.

A point of some interest in connexion with the development of Fisher's work is that while the first uses of the analysis of variance (e.g. in randomized block and Latin square designs) were being illustrated on data of field experiments in the early editions of *Statistical Methods for Research Workers* (1925, 1928), there was no parallel presentation of the underlying theoretical model. The paper (**12**, 1924) read before the International Mathematical Congress at Toronto, but not published till 1927, involved little more than a restatement of results that had been established for regression problems in (**6**, 1922), while the much later *Econometrica* paper (**13**, 1935) is only concerned with the properties of the χ^2, t and z distributions, and not with why the last distribution could be used in comparing component sums of squares in an analysis of variance derived from a particular experimental design. In more than one place in this volume, Fisher speaks somewhat scornfully of the mathematical statistician who attempted during the twenties and early thirties to teach statistics in the universities. He could not have realized the efforts involved in drawing up from the imperfectly presented outline of his ideas, proofs which could be accepted in the classroom. We do not know at what stage he himself realized the beautiful simplicity of his 'normal theory' analysis of variance model: the observations represented in multiple space by a point P belonging to a probability density cluster with hyperspherical contours; the problems of the experimenter expressible in terms of questions regarding the unknown co-ordinates of the centre O of that cluster; the art of analysis consisting in a cunning rotation of axes so that by a comparison of the squares of the projections of OP in separate co-ordinate spaces, the significance of 'effects' could be judged. Viewed from the geometrical angle, one 'sees' the result is true; one does not need an algebraic proof. But he did not, perhaps, give us very much help to see. A short section (pp. 97–8) in the paper dealing with the applications of 'Student's' distribution (*Metron*, 1925, **5**, no. 3), not reissued here, was no doubt written to establish this general idea, but its practical interpretation was not driven home convincingly. It was left to others, Irwin, Cochran, Kolodzieczyk, to attempt to translate the full consequences of the geometrical concept into more ordinary algebraic terms.

But there was perhaps another reason why no great effort was made to make clear the structure of this particular mathematical model. In the handling of agricultural experimental data it was soon realized that this construct, with its normal residual variation superimposed on a series of additive terms representing 'treatment' effects, which alone made the probability levels of the tests 'exact', often could not represent the true picture of experimental variation. Thus emphasis was placed on the need for randomization, and the reference set which gave meaning to the tests of significance became the randomization set and not the set that would be generated by resampling of the normal residuals. This second interpretation was justified practically, but it meant that the probability integrals of z or F are introduced into the tests as approximations only. Generally very good approximations no doubt, but still approximations whose accuracy has hardly yet been fully explored. On this extremely important point, the present volume has, however, scarcely a word to say from the mathematical angle. It was inevitable that many genuinely misunderstood what was in the minds of mathematicians at Rothamsted.

An interesting paper of a relatively early date is reprinted from the *Journal of the Ministry of Agriculture* (**17**, 1926); in this Fisher explains in simple terms the objectives that were being aimed at in the planning of experimental lay-outs. Printed just before this is the paper on tests of significance in harmonic analysis (**16**, 1929), in which he exercised his skill in hyperspace geometry to derive the distribution of the ratio of extreme value to the mean (or rather n times the mean) in samples of n from an exponential distribution. This paper has been completed by the addition of a new table giving the 5 and 1 % points for the ratio, for the range $n = 5$ to 50.

Two outstanding papers from the mathematical viewpoint belong to this middle period (**14**, 1928) and (**20**, 1929). In the former Fisher derived the sampling distribution, (A), of the multiple correlation coefficient R^2, in series form, and also two allied distributions, one, (B), which has been termed the distribution of a non-central χ^2 and a second (C) which is the distribution of what may be termed a non-central F, involving the ratio of a non-central to a central χ^2. It is likely that all three distributions have a larger

part to perform in statistical theory than they have so far played. In the second paper of 1928 he took up the problem of the sampling moments of moments and product moments. A new combinatorial technique, accompanying his introduction of the k-functions whose expectations were exactly the cumulants or semi-invariants κ, broke through what had become to appear an impassable barrier to further advance. The paper (**21**, 1930), giving the derivation of the moments of $g_1 = k_3/k_2^{\frac{3}{2}}$ and $g_2 = k_4/k_2^2$, provides an excellent example of the application of this technique.

The year 1930 saw the publication of the first of the four papers included in this series on fiducial probability (**22**, **25**, **27** and **35**). There have been one or two others, not included. The subject is one of those which has been surrounded with controversy and, as in other instances in Fisher's scientific career, this has undoubtedly arisen in part because many of his readers have been unable to grasp his meaning. This difficulty has arisen particularly over his use of the theory of probability. Thus in this 1930 paper there appears to be a clear statement that he regards as of first importance the relation between probability and frequency in repeated trials. Speaking (in **22**, p. 534) of the lower fiducial limit for the correlation coefficient ρ when the sample $r = 0{\cdot}99$, he writes: 'The value of ρ can then only be less than 0·765 in the event that r has exceeded its 95 % point, an event which is known to occur just once in 20 trials. In this sense ρ has a probability of just 1 in 20 of being less than 0·765.' Similar linking of probability and frequency in repeated trials could be found in many other places; for example, in his simple exposition of experimental design (**17**, 1926), he illustrates what he means by 5 % significance levels by describing how one might draw a line between the 25th and 26th largest deviation found in a series of 500 annual observations. Yet as far as can be gathered from the comments in the author's note accompanying the fourth paper of this group, his critics have been seriously at fault for supposing that from 1925 onwards he had regarded it as necessary to relate significance levels to frequency of occurrence in repeated sampling! Annoyance at criticism, though regrettable, can be easily understood, but it is surprising that the author who makes a point of the importance of aiding the student to avoid common misapprehensions should have been content to leave his theory of fiducial probability with this far from clear presentation (**35**, 1939) of the Fisher-Behrens test.

Although they are relatively accessible it is useful to have together Fisher's papers from the *Annals of Eugenics* on problems arising in connexion with discriminant analysis (**32**, 1936; **33**, 1938; **34**, 1940; and **36**, 1939). A paper less easily available to statisticians and all the more welcome in this series is the joint paper from the *Journal of Animal Ecology* (**43**, 1943) dealing with the relation between the number of species and the number of individuals in a random sample of an animal population.

A publication of this character, with its author's notes assessing the value of his work as it appears to him to-day, is of course bound to be self-revealing. He has thrown himself open to the public view. Many of the notes will be helpful to the student reader, as intended, but in one instance at least the comments only serve to perpetuate an occasion when reason has been overcome by anger and emotion. Most readers will regret the inclusion of the Note on Paper (**29**, 1937) which, if nothing more, shows in its last sentences a profound ignorance of Karl Pearson's character and, indeed, of his contemporaries.

The volume is produced in an unusual form. The decision to print the papers in their original size, which leaves wide margins for all papers except those from the *Philosophical Transactions of the Royal Society*, was a wise one. But it must be a matter of opinion whether it was necessary to reproduce photographically manuscript corrections of errors or crooked lines crossing out whole pages of text, presumably in the author's own hand, or to let the veteran of 60 apologize for the young man of 30's lack of balance in composing a paper. This is a style of presentation of a living author's scientific work which may not be to everyone's taste. Be that as it may, the main thing is to have these forty-two papers readily available, and for that and the life's work which lies behind them, we should all like to express our gratitude.

E. S. PEARSON

Probability Theory and its Applications. Volume I. By W. FELLER. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1950. Price 48*s*.

This is the first part of a comprehensive treatise on probability theory. It deals with probabilities defined on discrete sample spaces only, and the more difficult theory of probability in a continuous sample space is left to a second volume. By imposing such a severe restriction on subject-matter in the first volume the author has left himself free to write a most thorough and inspiring account of this part of the theory.

The first five chapters deal with the elementary combinatorial rules of discrete probability and their application to classical problems. Although most of this is not new, the discussion is elegant, stimulating and modern in flavour, for example in giving an interesting account of the difference between Bose-Einstein and Fermi-Dirac occupation problems.

Two further chapters deal with the Binomial and Poisson distributions and their approximation by the normal distribution function, which is introduced here solely as a mathematical function. Three more chapters deal with the laws of large numbers and more general types of distribution function. These give the best and clearest account of the laws of large numbers for discrete distributions that the reviewer has come across.

After chapter 11, which deals with generating functions, we come to the more original part of the book. This begins with two most interesting chapters on the theory of recurrent events, which has been largely developed by the author in order to apply it to the remaining part of the book, consisting of four chapters on random walk theory, Markov chains, and evolutive stochastic processes with a continuous time. Much of the theory of these subjects is here presented in a highly original way.

This book is a treatise on probability and not on statistics, and as such must be regarded as the most important written in modern times. However, no attempt is made to provide anything like a comprehensive bibliography. At the end of each chapter is a series of exercises, many of them very interesting in themselves. From these one can apparently deduce that professors at Cornell often park their cars where parking is forbidden, play a great deal of poker and bridge, and sometimes have difficulty in finding the right key for their front doors on arriving home. The style is clear and amusing and there seem to be remarkably few misprints.

A book of this order of importance deserves the closest scrutiny and the following criticisms on matters of detail are put forward for that reason. On p. 24 the word 'sample' is restricted to samples drawn in a specified order and this may confuse some readers. On p. 38 the author says that a certain hypothesis about a particular set of data is to be rejected on the ground that the sample would have had an extremely small probability if this hypothesis were true. This is certainly not a valid method of reasoning in the case here presented, in which it is easy to construct examples in which the probability of any event is extremely small. For example we cannot reject the hypothesis that all permutations of a pack of 52 cards are equally likely because the probability of a particular observed order is about 10^{-68}. Similarly the statement on p. 56 that with a 1% level of significance we shall find assignable causes 99 out of 100 times is also clearly untrue. On p. 57 the statement that only the one-dimensional theory of runs has been considered ignores work done by Krishna Iyer and others. In the footnote on p. 120 it is not clear why only one degree of freedom is used in the χ^2 considered, and on p. 210 there is a misprint in example 7.

P. A. P. MORAN

Experimental Designs. By WILLIAM G. COCHRAN and GERTRUDE M. COX. ix + 454 pp. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1950. Price 46*s*.

Analysis and Design of Experiments. By H. B. MANN. x + 198 pp. New York: Dover Publications. 1950. Price $2.95.

These two books have similar titles and they both deal with the theory and practice of the wide range of statistical methods included under the general heading of 'analysis of variance'. Apart from this they have little in common. The difference is only mildly described by saying that the first book is more concerned with practice and the second more concerned with theory. The title of the book by Cochran and Cox is a fair indication of the contents, but it is to be wished that Prof. Mann had given a more accurate indication of the contents of his book in the title he assigned to it. An essential phrase in this title should be 'mathematical theory of', or words to the same effect.

Experimental Designs should prove especially valuable to statisticians requiring a compendium of all the more useful designs accompanied by an outline of the relevant theory, and also to experimenters with sufficient time, inclination and ability to apply modern statistical methods in a serious manner to the design, analysis and interpretation of their work. The first three chapters contain a brief introduction to tests of significance, estimation, and related concepts, followed by an account of the type of procedure employed in the analysis of variance. A warning must be given that the reader cannot expect to appreciate this part of the book without previous study of general statistical theory and some experience in the application of at any rate the simpler forms of analysis of variance. The third chapter, in particular, while an excellent summary of the principles underlying the analysis of variance, will not be fully understood at a first reading, without a background built from experience of the problems arising in a variety of experimental situations.

The next twelve chapters deal with different forms of experimental design. Each chapter contains a description of the conditions under which the particular design is likely to be useful, and the appropriate analysis, with a simple account of its theoretical basis. In the more complicated cases a selection

of suitable designs ('Plans') is given at the end of the chapter. This feature greatly increases the value of the work as a reference book.

The sixteenth chapter ('Random permutations of 9 and 16 Numbers') constitutes a rather odd, though useful, tailpiece.

It is claimed that *Analysis and Design of Experiments* is 'invaluable to three groups: mathematicians, students and teachers, practical experimenters and statisticians.' One cannot help feeling that this claim becomes less reasonable on passing through these groups and is definitely over-optimistic in regard to the later classes. It should be clearly understood that this is a mathematical text-book written in an advanced and concise style. This being accepted, however, it can be said that the author has produced a very good book of this kind, which would be even better without the occasional brief and rather lonely references to 'practical applications'.

Five pages suffice to cover the fundamental results on the χ^2 and F distributions; a further ten pages cover 'the multivariate Normal distributions' and the distributions of quadratic forms. Chapters 3–6 deal with general analysis of variance theory, including the approach via the likelihood ratio test applied to the general linear hypothesis. The next three chapters cover orthogonal Latin squares (treated at relatively great length) and the construction and analysis of incomplete block designs. There follow chapters on non-orthogonal data, factorial experiments, randomized blocks and quasi-factorial designs, analysis of covariance (two pages!) and interblock estimates and interblock variance. In addition to the usual F-tables, the author includes Tang's tables (*Statistical Research Memoirs*, **2**, 126), from which the power function of the analysis of variance tests can be evaluated in certain cases.

Despite its austere style and the inequalities of treatment of the various parts of the subject, this book is a useful summary of the modern mathematical theory which has grown up around the analysis of variance.

N. L. JOHNSON

First Course in Probability and Statistics. By J. NEYMAN. New York: Henry Holt and Co. Inc. 1950. Price 30*s*.

As indicated in the Preface, this book follows the lines of a course of lectures given mainly to beginners in mathematical statistics at the University of California. Thus it is assumed that the reader's mathematical knowledge is limited to 'high-school algebra', and to avoid the use of calculus the problems tackled are those in which the variable concerned has a discontinuous distribution. What the author has termed the fundamental probability set remains, therefore, finite, and theorems in probability can be derived rigorously without making an appeal to the theory of sets. But while the mathematics can in this sense be kept simple, the author aims at introducing from the beginning what he regards as the basic concepts of statistical theory and connecting these concepts with various fields of application. Certain starred sections carry the subject rather further for graduates who, while beginners at statistics, will in general be of substantially higher intellectual maturity.

The special class of students in view has no doubt dictated to some extent the line of approach to the subject, but as in the case of any course of lectures given by a teacher of outstanding individuality, the structure of the book reflects strongly his own outlook on the subject. For this reason it is particularly interesting to have set out in simple form what may be termed Prof. Neyman's philosophy of statistical thought. The characteristic feature of his contribution to statistics during the past 25 years has lain in his conviction that the more empirical British and American approach to mathematical statistics needed strengthening by a closer and more sure integration with the theory of probability. This need arises not only in the development of the mathematical theory of distributions, but more importantly in defining the relation between our calculus and the way we think.

In a paper of 1937 Neyman suggested that we might properly describe the adjustment of our action to a limited amount of observation as 'inductive behaviour'. In ordinary affairs of life the conscious part of this adjustment is based on certain rules—when I see this happening, then I do that. Inevitably we judge the value of the rules by the consequences which result from following them. This general concept may be translated into the mathematical field of probability and statistics. When the observations with which we are faced are of a random or probabilistic nature, the study of these rules of inductive behaviour becomes the concern of the mathematical theory of statistics. This uses the theory of probability to derive the performance characteristics of the rules and, in particular, to study how to determine rules which have certain desirable or optimum properties. Thus mathematical statistics is a branch of the theory of probability. The basic ideas, now developed to a more formal stage, are those which emerged from the association of the author and the reviewer during the years 1927–33.

The plan of the book is as follows. Chapter 1 is of an introductory character, indicating where the systematic instruction in probability theory of the next three chapters is leading. Chapter 2 deals with the

fundamental concepts and theorems of probability; a long section on 'competing risks' illustrates how these concepts may be used to construct a mathematical model which could be employed in studying the effectiveness of a specified treatment of a recurrent disease. Chapter 3 shows how probability models underlie the theory of genetics. That the models here used may not be adequate, through inevitable simplification, to describe observed phenomena is perhaps not material to the author's theme. For, he writes, 'since the main purpose of discussing genetics is to provide interesting illustrations of the application of probability, it is natural to introduce limitations'. In Chapter 4 the concept of a random variable is introduced and the binomial and hypergeometric distributions are derived and discussed. Since the book confines attention to finite sets, the normal distribution is only considered as providing a means of approximation to the sum of binomial terms.

Finally, with the student thoroughly grounded in probability theory, the volume is completed by a 90-page section (Chapter 5) on 'Elements of the theory of testing statistical hypotheses'. This provides the main purpose of the book. Within the limits imposed, a simple and precise exposition is given. The argument is developed stage by stage, through a series of definitions whose meanings are illustrated by examples and driven home by problems and exercises left for the student to solve. Considerable space is given to a discussion of R. A. Fisher's problem of the 'lady tea-taster', and considerations involved in the choice of a critical region are clearly illustrated on a bivariate problem, such as might arise in using two tests in a screening procedure for tuberculosis. The concepts of a best critical region and of a uniformly most powerful test emerge simply and naturally; the chapter ends with a brief discussion of the likelihood ratio or λ-principle as a means of determining an appropriate rest.

The style of presentation is frankly didactic, with ideas driven home by repetition in the exercises. But this must be expected, since one object of the book is to get across to first-year College boys ideas that have more commonly been reserved for a later stage of instruction.

In his introductory chapter the author has suggested that a common pattern can be traced in the development of different branches of mathematics. There is first a period in which certain 'permanencies' are observed among natural phenomena and these create a number of problems in inductive behaviour. Next, the mathematician steps in and attempts to form abstract models of the phenomena, elevating the rather vaguely appreciated permanencies to the level of basic concepts and axioms, testing the adequacy of the models and giving greater precision to the rules of behaviour. Finally, a stage is reached where most of the mathematical problems considered are dictated by pure curiosity as to the consequences of the axioms assumed and their mutual relations, one with another; the mathematical discipline becomes then a branch of pure mathematics.

Neyman, himself, may be described as standing between the second and third stage of this picture. He is intensely fascinated by the construction of his models, but through his many contacts with research establishments endeavours to keep these models in gear both with what the experimentalist observes and with the processes of thought which guide his behaviour. Nevertheless, it seems likely that a considerable proportion of the students whose first course in statistics is based on this book as text, will end by regarding statistics as a branch of pure mathematics, and in their own later research they may pay only lip service to the practical problems which the theory might be directed to solve.

No doubt this dividing of ways in inevitable, but if so, it seems to follow that even at an initial stage, courses in applied and in pure mathematical statistics must be treated separately. The basic concepts which Neyman has at heart can be introduced in several ways at an elementary level, but the drill required in training the applied statistician must surely involve more handling of numerical data, and less practice in the formal presentation of simplified problems in a way to meet the requirements of a rigorous mathematical discipline.

E. S. PEARSON

Sampling Methods for Censuses and Surveys. By F. YATES. Pp. xiv + 318. London: Charles Griffin and Co. Price 24*s*.

This is an excellent book.

It is well known that for many purposes the selection of a sample from a population can give under favourable circumstances almost as much information as taking a complete census—in fact, in some cases the sampling method has been found more accurate, when there has been difficulty in obtaining a complete and reliable census return. This book gives a detailed discussion of the practical problems involved in planning and executing such a sample survey, the methods of extracting information from the data once it has been collected, and the determination of the accuracy of the final results. All the necessary formulae used in computations are set out and carefully explained: but the author has not given the theoretical derivation of these formulae (no doubt feeling that this book is not really a suitable place for the inclusion of a considerable amount of mathematical statistics).

Chapter 1 discusses the relative advantages of sampling and complete census methods. Chapter 2 shows how to avoid biased and unreliable samples (with illustrations pointing out the errors which can occur as the result of bad selection). Chapter 3 describes the various types of sample—the completely random sample, useful for a fairly homogeneous population, and the various kinds of stratified samples which can be used to increase accuracy when the population to be studied contains units of very different sizes. Chapters 4 and 5 deal with the planning, execution, and analysis of the survey, including such questions as the exact determination of the units to be sampled, the instructions to be given to the field investigators, and the recording of the results on punched cards. Chapter 6 deals with the estimation from the sample of the means and totals of the population values, and chapter 7 the estimation of the error in these estimates, using various methods of sampling. In the last chapter the question of 'efficiency' of a sample survey is discussed. By performing a small 'pilot' survey, or by using experience gained in previous surveys, an estimate can be obtained of the efficiency of different methods of sampling the population. This enables the survey to be designed in such a way as to cost as little as possible for a given accuracy. Sometimes when the information obtained in a survey can be considered as of direct financial value, it may be possible to balance the cost of obtaining extra information against the benefits obtained from greater accuracy, and so obtain the most favourable result.

The style throughout the book is very lucid, and no mathematical knowledge is assumed beyond simple algebra and the use of summation signs. Dr Yates makes an ingenious use of different styles of type to distinguish population values and estimates, which makes the formulae easy to read and comprehend (though less easy to distinguish when spoken). Wherever it is suitable, the methods are illustrated by worked examples. The index has been carefully prepared. There are a number of helpful practical hints scattered throughout the book: and, to sum up, the book is one which can be cordially recommended to all concerned in the planning and execution of surveys.

CEDRIC A. B. SMITH

Symposium on Stochastic Processes. 'Stochastic processes and statistical physics.' By J. E. MOYAL. 'Some evolutionary stochastic processes.' By M. S. BARTLETT. 'Stochastic processes and population growth.' By D. G. KENDALL. Published in the *Journal of the Royal Statistical Society*, Series B, Vol. XI, 1949, pp. 150–282. Cost of reprint 12*s.* 6*d.* post free.

This symposium marks a break from the hitherto intensive pre-occupation with stationary series, with the accompanying concentration on correlogram and periodogram analysis. It is evident after reading the three papers presented, if not before, that there are unending fields of application for more general models. It is true that most of the situations considered are of a fairly precise physical nature, where one has strong *a priori* reasons for expecting a particular mathematical model to prevail. That is, the reasoning is direct rather than inverse. Even so, the problems arising are deep ones, and the methods here developed go a long way towards their solution.

Mr Moyal treats several problems of mathematical physics, notably of quantum mechanics, where a stochastic element is present. It may be present only in the initial conditions; given these, the system develops deterministically. This is a 'crypto-deterministic' system. On the other hand, new random elements may be continually entering the system as in the Brownian motion, and this is the case of chief interest.

The author's presentation of his 'theory of random functions' is a fine piece of work. The material is necessarily rather condensed, but it shows almost throughout a unity and continuity of treatment which lighten the reading considerably. It is instructive to see the strict parallel between a calculus of random functions and a calculus of ordinary determinate functions. Modes of convergence, properties such as differentiability and analycity all have their direct analogues.

On the other hand, the abrupt hop from an analytic treatment to the use of abstract spaces would be rather disconcerting to one who had not previously read Kolmogoroff's and Karhunen's own works.

Moyal defines a random function initially by giving the simultaneous distribution function of the dependent variable, 'x', for all arbitrary finite sets of values of the independent variable 't'. As he himself points out, such a definition is often insufficient if we wish to consider the behaviour of 'x' over continuous 't' intervals, and in this connexion some interesting unpublished work due to Prof. Pitt is outlined. To the applied worker, this is, of course, a finesse.

The author gives a definition of the 'degree of indeterminacy' of a random function in terms of the number of functional relationships required to determine the whole course of the process. This is clearly akin to H. Wold's 'rank of singularity' (1938), although Moyal does not limit himself to the case where the functional relationships are linear.

Moyal strongly emphasizes the importance of the auto-covariance function, whose analytical properties largely determine those of the process. This is clearly seen in the sections on stochastic convergence, differentiability, and the existence of the Riemann and Lebesgue integrals of the process.

Among examples of the direct establishment of stochastic equations, the author gives the Brownian motion and the shot effect. The appropriate relations here are often written as usual differential equations, although the derivative of highest order does not exist. To avoid this Moyal uses two distinct differential operators w.r.t. time: $\delta/\delta t$ and d/dt. This seems as convenient a method as any.

The very general treatment of the behaviour of an arbitrary stable linear system into which a stochastic impulse function is being injected is interesting, particularly in view of the prediction theory for such systems recently developed by Dolph & Woodbury (unpublished).

The section on the deduction of equations giving the time behaviour of the distribution or characteristic function is of great interest. This is largely related to M. S. Bartlett's work. Equation 8·1·8 is of great generality, as it is obtained by substituting the limiting form of the 'transition' probability in the general consistency relation, and merely assumes that this limiting transition probability exists. It is tempting to surmise what this equation has for interpretation. If only the first derivate moment is non-zero the equation is the continuity equation of a fictitious 'probability fluid'. The more general equation has also something of this character: if the variate increment in an infinitesimal time increment has a probability of a fixed order of magnitude, then the differential coefficients of the probability are to some extent restricted, so that spontaneous discontinuities do not occur.

Moyal gives ample illustration of the advantages of the method in applications. For the Langevin equation with one degree of freedom, the very generality of the method makes a direct solution preferable, but its virtues are apparent in the case of several degrees of freedom.

Quite apart from the results of his own researches there presented, Moyal's paper is most impressive in the way it so firmly and naturally unites several branches of science. For the reviewer, at least, the reading of the article cast a new light upon the subject of stochastic processes and its scope.

Prof. Bartlett chose to deal with some evolutionary processes, and, in the spirit of the symposium, gave an advanced treatment of several problems of genuine and urgent practical interest. His paper fell into two sections, each of which centred about a particular mathematical relation.

For the first section, on additive processes, his relation was Wald's identity for the number of random variates whose sum is to fall between fixed limits. Letting one limit degenerate to minus infinity, the author obtains an upper bound for the probability that the other limit is some time overshot. As the author briefly indicates, the result is also valid for the continuous case. A. Khintchine's well-known monograph (1933) contains much that is relevant here, as it does also for the diffusion equation below.

The asymptotic results obtained on the basis of this identity for the insurance and renewal problems have, of course, been reached before by direct methods. This does not rob the demonstration of its interest, however, and it is plain to see that a whole class of asymptotic results could be thus obtained, with very little effort.

The application of the theory of additive processes to obtain a distribution of Kolmogoroff's was a surprising one. Surprising, because it is by no means immediately seen that there is any connexion between this problem of deviations of an empirical distribution curve from theory, and stochastic processes. However, Prof. Bartlett's rather condensed reasoning soon shows that there is, and the result follows delightfully simply. The problem is reduced to the consideration of the diffusion equation with given boundary conditions, and it is probable that there are related problems which could well be treated in this manner.

The second section, on multiplicative processes and extensions of these, is built upon the relation derived from the consistency condition and the limiting form of the transitional probability, to which we have already drawn attention in Moyal's paper. The equation is most used here as a relation between probability generating functions. The author concerns himself largely with the multivariate case, corresponding to several 'states', and it is soon apparent that the equation for a given Markov model may be written down almost directly, by following a few simple rules. One can say that the methods presented here virtually give an immediate solution to any problem of a Markov nature, if we mean by a solution that all probability considerations have been reduced to definite mathematical relations; in this case, certain partial differential equations. However, these equations were explicitly soluble only in two cases, one of them being the case of a conservative system when the equation was linear with linear coefficients. Otherwise, the way was blocked by the equations' non-solubility.

The real need here seems to be a general treatment of the type of equation arising. That this will not be simple is shown by the equations' intractability even in the deterministic case. Bartlett notes that the first few terms of a power series expansion gave a good approximation to the empirical results in one deterministic case, this suggests the feasibility of investigating series solutions of some sort for

the stochastic models also. An alternative is that a relatively large group of the equations and initial conditions encountered may ultimately involve only one or two functions not already tabulated. Tabulation of these would give the models some practical value.

It is quite obvious that the construction of an equation in the characteristic function gives a method of great generality and power. Bartlett's paper solves many problems and reveals yet more which remain to be solved.

Mr Kendall's paper gives an almost complete exposition of population growth mathematics, from the very simplest deterministic model up to his own interesting treatment of more realistic explanations. Here, as in Bartlett's paper, it is evident that we are now in a position to set up equations for models of almost any degree of refinement. The difficulty is to solve them. Many of these equations, Kendall's (111) for example, are of new types and would repay investigation for their own sake.

The author uses a variety of techniques: direct attacks and equations in the characteristic or probability generating functions. His use of functionals of an arbitrary function to treat a continuous age distribution is something which opens up new fields. He himself points out that the solutions have the drawback of a discontinuity for an age equal to current time. Nevertheless, this seems a very natural method of attacking the problem, and the difficulty could perhaps be overcome. After all, the discontinuity is not a meaningless one.

It seems probable that Kendall's 'multiple phase' process, which he develops in connexion with bacterial colonies, could be adapted to the case of an animal population with a birth-rate dependent upon age. This could be achieved by permitting reproduction in several phases with a probability intensity appropriate to the phase. Such a treatment would, of course, introduce the complication of two distinct ages, a temporal age and a 'virtual' age, determined by the phase reached. An effect which is, perhaps, realistic enough.

Kendall emphasizes strongly an effect observed by Feller, that the mean behaviour of a stochastic model does not necessarily coincide with the behaviour of the corresponding deterministic model. His treatment of the logistic case illustrates convincingly this rather unexpected fact.

The result for the estimate of the birth-rate in a purely reproductive process is a neat one. We may expect that the estimation problems will be formidable for more complicated models, but, as Kendall observes, the difficulty does not arise until the models themselves have been found worthy. The estimate variance is here based upon the supposition of a large number of replicates. This supposition may be fulfilled in the case of bacterial or insect populations, but for human populations, at least, we will rather possess a relatively long set of observations upon the same material. The estimate variance must thus be derived rather differently. We can, of course, split the population into sub-populations, but whether these develop independently is another matter.

The discussion following the presentation of the three papers amply attested to their value and to the stimulation which is felt even with a reading of them. The best possible commentary upon such a symposium would be the development under the next year or two of material for a third symposium, and there is little doubt but that this will be realized.

P. WHITTLE

University Institute of Statistics, Uppsala

REFERENCES

DOLPH, C. L. & WOODBURY, M. A., University of Michigan. Unpublished memorandum.

KHINTCHINE, A. (1933). Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. *Ergebn. Math.* **4**, no. 3.

WOLD, H. (1938). A Study in the Analysis of Stationary Time Series. Dissertation, Stockholm.

CORRIGENDA

E. S. PEARSON, *Biometrika* (1950), **37**

p. 386. A 2 has been omitted outside the brackets in equation (12), which should read:

$$\bar{z} = \int_0^1 z\,du = 2 - 2\{F(X)\log_e F(X) - F(X-1)\log_e F(X-1)\}/f(X).$$

(*All Rights reserved*)

BIOMETRIKA. Vol. 38, Parts 1 and 2

CONTENTS

PAGE

Major Greenwood (*With Frontispiece*) 1

Tables of the 5% and 0·5% points of Pearson curves (with argument β_1 and β_2) expressed in standard measure. By E. S. Pearson and Maxine Merrington 4

Regression, structure and functional relationship. Part I. By M. G. Kendall 11

Partial and multiple rank correlation. By P. A. P. Moran 26

An application of the distribution of the ranking concordance coefficient. By A. Stuart 33

The effect of non-normality on the power function of the F-test in the analysis of variance. By F. N. David and N. L. Johnson 43

Efficiency of the method of moments and the Gram-Charlier Type A distribution. By L. R. Shenton 58

On distributions for which the Hartley-Khamis solution of the moment-problem is exact. By H. P. Mulholland 74

Estimation problems when a simple type of heterogeneity is present in the sample. By W. M. Long 90

Some tests for randomness in plant populations. By Marjorie Thomas 102

Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. By E. S. Pearson and H. O. Hartley 112

Some questions of distribution in the theory of rank correlation. By S. T. David, M. G. Kendall and A. Stuart 131

Note on an exact treatment of contingency, goodness of fit and other problems of significance. By G. H. Freeman and J. H. Halton 141

The geometry of estimation. By J. Durbin and M. G. Kendall 150

Testing for serial correlation in least squares regression. II. By J. Durbin and G. S. Watson . . 159

Bi-variate k-statistics and cumulants of their joint sampling distribution. By M. B. Cook . . 179

Random dispersal in theoretical populations. By J. G. Skellam 196

The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. By A. K. Gayen 219

Miscellanea

Some observations on the practical aspects of weighing designs. By K. S. Banerjee . . . 248

Test for the significance of the difference between means in two normal populations having unequal variances. By D. G. C. Gronow 252

Reviews

R. A. Fisher's 'Contributions to Mathematical Statistics' 257

W. Feller's 'Probability Theory and its Applications', vol. I 259

William G. Cochran and Gertrude M. Cox's 'Experimental Designs' 260

H. B. Mann's 'Analysis and Design of Experiments' 260

J. Neyman's 'First Course in Probability and Statistics' 261

F. Yates's 'Sampling Methods for Censuses and Surveys' 262

'Symposium on Stochastic Processes' 263

A volume of *Biometrika* containing about 400 pages, with plates and tables, will be published annually in two half-yearly issues.

Papers for publication should either be sent to

PROFESSOR E. S. PEARSON, Department of Statistics, University College, London, W.C. 1,

or if more convenient to

Dr John Wishart, Statistical Laboratory, St Andrew's Hill, Cambridge.
Professor M. G. Kendall, London School of Economics, Houghton St., London, W.C. 2

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 25 copies of their papers free. Joint authors 15 copies each.

The subscription price, *payable in advance*, is Inland 45*s. net* per volume and Abroad 54*s. net* (including packing and postage). At present certain early volumes are out of print, but where available the cost is Vols. 1 to 32, £5 each, and Vols. 33 to 36, 63*s.* each (excluding postage). Index to Vols. 1 to 5, 2*s. net*. Index to Vols. 1 to 15, 5*s. net*. Bound volumes, of those volumes in print, are again available at a charge of 10*s.* extra per volume; also binding cases at 4*s.* 6*d.* each. Cheques must be made payable to *Biometrika*, crossed "*a/c Biometrika Trust*" and sent to The Secretary, *Biometrika* Office, Department of Statistics, University College, London, W.C. 1, to whom all orders for series, single copies and offprints should be addressed. All foreign cheques must be drawn in sterling and on a Bank having a London Agency.

First Printed in Great Britain at the University Press, Cambridge
Reprinted by offset-litho by Percy Lund Humphries & Co., Ltd., Bradford

PAGE
1

4
11
26
33

43
58

74
90
102

112

131

141
150
159
179
196

219

248

252

257
259
260
260
261
262
263

o half-

not be

ng and
d Vols.
mes, of
Cheques
artment
dressed.